

RAWLSIAN ETHICAL ACT CONTRACTARIANISM

---

A Dissertation Presented to the Faculty of the Graduate School  
University of Missouri

---

In Partial Fulfillment  
of the Requirements for the Degree

Doctor of Philosophy

---

by  
JOEL DITTMER

Dr. Brian Kierland, Dissertation Supervisor

MAY 2010

The undersigned, appointed by the dean of the Graduate School, have examined the  
dissertation entitled

RAWLSIAN ETHICAL ACT CONTRACTARIANISM

presented by Joel Dittmer,

a candidate for the degree of

doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

---

Professor Brian Kierland

---

Professor Peter Vallentyne

---

Professor Robert Johnson

---

Professor Paul Weirich

---

Professor Ann Bettencourt

## ACKNOWLEDGMENTS

In my case, I owe a lot to a number of people in receiving a Ph.D. Let me start with acknowledging my doctoral committee. I want to give great thanks to Professor Brian Kierland (my advisor), Professor Peter Vallentyne (my second reader), Professor Paul Weirich, Professor Robert Johnson, and Professor Ann Bettencourt (my outsider reader). I'll start with Robert. Thank you so much for being my first philosophy professor, an introductory course in ethics; I loved this course, and it hooked me. Also, thank you for being my advisor for my master's thesis. Thank you Paul. I took two courses with you. I did well in both, and in both, you encouraged me, in particular the seminar on epistemic logic. Professor Bettencourt, I genuinely appreciate your service on my committees for the master's, the doctoral comprehensive exams, and the dissertation. Given that you are the chair of the Psychology Department, I am more than grateful for you taking time to take my professional development seriously. As for Peter, thank you so much! You've really helped me to focus my thoughts, and in getting me to convey my thoughts in a comprehensible way. In addition, thanks to your guidance, I know how to motivate a piece of work and keep in mind the bigger picture.

I have to express great appreciation to my advisor Brian Kierland. Brian has guided my training throughout my graduate school experience. Between seminar papers, a published article, my master's thesis, my dissertation, and job preparation, I estimate somewhere around, or perhaps well over, a 1000 comments by Brian on my work. Throughout all of these comments, as well as office meetings, lunch meetings, and phone

conversations (some of them up to three hours!), he has invested an extraordinary amount of time and energy in my development. Thank you Brian.

Finally, I want to thank all of my colleague graduate students; it's been a blast! In particular, though, I should thank: Adam Carter, Alan Tomhave, Bill Ramey, Dan Marshall, Devin Frank, Eric Roark, Jason Hedderman, Jeff White, Jon Trerise, Michael Hartsock, Omar Moad, and Tony Thomas.

Also, I have to, in an extremely understated way, thank Mom, Dad, and Beeker.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	ii
ABSTRACT.....	v
Chapter	
1. INTRODUCTION: A NEW CONTRACTARIAN ETHICAL THEORY.....	1
2. NON-RAWLSIAN ETHICAL CONTRTACTARIANISM.....	12
3. ACT VERSUS RULE CONTRACTARIANISM .....	49
4. RAWLSIAN ETHICAL ACT CONTRACTARIANISM AND THE VEIL OF IGNORANCE.....	72
5. CONCLUSION.....	110
BIBLIOGRAPHY.....	116
VITA.....	119

## ABSTRACT

Assuming that contractarianism is appropriate for developing an ethical theory, which contractarian ethical theory is best? My dissertation provides an answer to this question. Drawing on the work of Rawls, I provide an ethical theory which most importantly employs a veil of ignorance. My account contrasts with both Gauthier's Hobbesian style ethical theory as well as with Scanlon's Kantian-inspired theory. As such, I explain the deficiencies of each account with respect to certain problems typically endemic to contractarianism. Because there have been various objections to contractarian accounts employing a veil of ignorance, I argue that these objections pose no serious problem for my account. Another feature of my account is that it is an act-based theory opposed to a rule-based one. Typically, contractarianism has been formulated as a view in which contracting agents agree upon rules (or: principles), and it is with these rules that actions are then evaluated as morally permissible or impermissible. In contrast to the typical rule-based view, I contend that the contracting agents agree upon actions.

## CHAPTER 1 – INTRODUCTION: A NEW CONTRACTARIAN ETHICAL THEORY

The most famous contemporary contractarian theory is proposed by Rawls in his *A Theory of Justice*. His theory can be characterized as follows:

Rawls' contractarianism: *X* is just if and only if it accords with the principles that the contracting agents would agree to when placed in the original contracting situation, where this amounts to the contracting agents being rational, self-interested, mutually unconcerned, as well as having limited information under the constraint of being placed behind a veil of ignorance.

Here, *X* refers to political institutions and their policies. Put this way, Rawls' theory is then a theory of justice. I'm going to argue for it (or: a modified version of it) as an ethical theory, where the contracting agents do not evaluate the justice of political institutions and their policies, but instead the permissibility actions of individual agents. In proposing Rawls' theory as an ethical theory, it is appropriate that I call it a kind of Rawlsian ethical contractarianism.

Another feature of my account, which I will explain in more detail later in this chapter, is it being a direct theory opposed to an indirect one. Instead of the objects of agreement being principles, or rules, by which actions are then evaluated, the direct objects of agreement are evaluations of actions. Following the act versus rule distinction with respect to utilitarianism (act is direct, rule is indirect), my account is a kind of act contractarianism. The Rawlsian and the act components together form a view which can be appropriately called "Rawlsian ethical act contractarianism."

The question, then, that I'm answering in this dissertation is this: For those who think that contractarianism is the correct approach in developing an ethical theory, why is the Rawlsian ethical act contractarianism the best approach? In answering this question, a natural place to start

is by giving a sketch of the various contending contractarian ethical theories. After doing this, I will conclude this chapter with (i) a brief characterization of my own view and (ii) a sketch of how the subsequent chapters proceed.

## 1.1 OTHER CONTRACTARIAN ETHICAL THEORIES

### 1.1.1 Background concerning contractarianism

Although this dissertation is not a work in the history of philosophy, it's important to understand in general how the proposed approach differs from traditional contractarian accounts. Traditional contractarians, such as Hobbes, Locke, and Rosseau each provide a kind of contractarianism that differs largely from the one that I develop.<sup>1</sup> First, their contractarianism is very much concerned with the legitimacy of political authority (i.e., obligation to obey the law).<sup>2</sup> Perhaps they can be read as developing theories of justice just as Rawls does with his contractarianism. I, on the other hand, am not proposing contractarianism as a theory of political obligation or of justice. Instead, I'm proposing a kind of contractarianism as an ethical theory – a theory that fills in the right-hand side to “An action is morally permissible if and only if ...”. Second, the contractarianism of the past, as well as of the present, has construed the objects of agreement between the contracting parties to be rules, or principles.<sup>3</sup> Consider Rawls' two main principles of justice as the object of agreement. I challenge this idea, not because I think those are the wrong principles per se, but instead on the grounds that the object of agreement for the contracting agents are not principles. I

---

<sup>1</sup> See John Rawls. (1971) *A Theory of Justice*. On p.11, fn.4, Rawls writes: “As the text suggests, I shall regard Locke's *Second Treatise of Government*, Rousseau's *The Social Contract*, and Kant's ethical works beginning with *The Foundations of the Metaphysics of Morals* as definitive of the contract tradition. For all its greatness, Hobbes' *Leviathan* raises special problems.” The idea that Kant is a contractarian is controversial (although Derek Parfit supports the view that Kant is a contractarian in his forthcoming *Climbing the Mountain*), and that Hobbes is not, seems to me, even more controversial. Gregory Kavka's *Hobbesian Moral and Political Philosophy* and Jean Hampton's *Hobbes and the Social Contract Tradition* are two recent works that do not argue per se that Hobbes is a contractarian, but instead assume that he is, so to make further claims concerning the validity of his contractarian argument. Nevertheless, cases can be made for the idea that Hobbes is not a contractarian.

<sup>2</sup> For an interesting application of contractarianism to argue for a sort of libertarianism, see Jan Narveson. *The Libertarian Idea*. (1988).

<sup>3</sup> In commenting on Locke, Rousseau, and Kant as the established figures of contractarianism, Rawls writes that “the guiding idea is that the principles of justice for the basic structure of society are the objects of the original agreement” (Rawls, p. 11).

characterize and motivate a kind of contractarianism where the objects of agreement are decisions about actions in specific choice situations.

*1.1.2 Three features of Rawls' theory, and Rawls' thoughts on how it can be used as an ethical theory*

There are three things to distinguish with respect to Rawls' theory. There is (i) Rawls' theory of justice as justice of fairness, (ii) Rawls' two principles of justice, and (iii) the argument that the two principles of justice are the content of justice as fairness. Justice as fairness is something like this: X is just if and only if it accords with the principles that the contracting agents would agree to when placed in the original contracting situation, with such and such characteristics. The argument that the two principles are the content of justice as fairness would just amount to showing that the principles the contracting agents would agree to are the two principles. And the two principles are as follows:

First Principle: Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.

Second Principle: Social and economic inequalities are to be arranged so that they are both: (a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and (b) attached to offices and positions open to all under conditions of fair equality of opportunity.<sup>4</sup>

These principles, although interesting and worthy of much discussion, are not what I'm interested in defending in giving a Rawlsian ethical theory. Of the three distinguishing features of Rawls' account, I'm not interested in defending his two principles, and so, I'm not interested in defending the argument from his fundamental theory of justice as fairness to these two principles. Instead, I'm interested in defending, mostly, his fundamental theory as applicable for an ethical theory (not, as applicable for a theory of justice).

---

<sup>4</sup> Rawls 1971, p.302. I should note that in addition to these principles, Rawls thinks that there is a priority rule which orders the two principles. This rule specifies that the first principle takes priority.

Although Rawls offers his fundamental contractarian theory as a theory of justice, he does say some things with respect to how his contractarianism relates to ethical theory. He considers principles that apply to individuals, which divide into two main groups – requirements and permissions. Requirement principles are further classified into two kinds – obligations and natural duties.<sup>5</sup> As for examples, Rawls cites obligations of fairness and of fidelity. For natural duties, Rawls mentions principles of upholding justice, of mutual aid, of mutual respect, of non-injury, of not-harming innocents.<sup>6</sup>

Rawls writes: “It may be observed, though, that once all the principles defining requirements are chosen, no further acknowledgments are necessary to define permissions.”<sup>7</sup> The most plausible rendering of what Rawls is saying here is the following: All principles of moral requirement are decided from the original position – thereby defining all principles of permissibility – and yet such principles will still need to be applied to specific moral cases. This point about all principles, whether they be of requirement or of permissibility, invites the following question: How exactly are Rawls’ principles – such as his principles of mutual aid, of mutual respect, of non-injury, of not-harming innocents – to be applied to specific moral cases? More specifically, how are these principles to be applied to cases of abortion, euthanasia, genetic research, famine relief, treatment of terrorist suspects, property disputes, as well as idealized moral cases like the trolley case and the organ-transplant case? As I will explain later in the dissertation, it’s partly because I don’t think that Rawlsian principles will be sufficiently

---

<sup>5</sup> In clarifying the distinction between the two, as well as understanding how they are generated, we can take notice of the following passage by Rawls:

Thus even though the principles of natural duty are derived from a contractarian point of view, they do not presuppose an act of consent, explicit or tacit, or indeed any voluntary act, in order to apply. The principles that hold for individuals, just as the principles for institutions, are those that would be acknowledged in the original position. (Rawls 1971, p.302)

(I should note that in addition to these principles, Rawls argues for a priority rule which says that “The principles of justice are to be ranked in lexical order and therefore liberty can be restricted only for the sake of liberty.”) So although principles of natural duty and of obligation are recognized by the contracting agents, that someone has a specific obligation requires some sort of voluntary act on their part, whereas natural duties apply to all persons unconditionally, independently of whether they’ve performed some voluntary act. In fact, Rawls calls principles of natural duty “unconditional principles” – (Rawls, p.116).

<sup>6</sup> Rawls, p. 109.

<sup>7</sup> Ibid., p. 230.

applicable to specific moral cases that a main difference between my account and his is that I reject the idea that principles, or rules, are the objects of agreement by the contracting agents.

### *1.1.3 Richards' Rawlsian contractarian ethical theory*

Another contractarian ethical theory to consider is one developed by David Richards. Richards' account is very much like Rawls'. Unlike me, he uses a Rawlsian account to develop a comprehensive theory of morality – one that covers both the laws and policies of institutions and the actions of individuals. So it covers both matters of justice and matters of morality (in the narrower sense). The following is a broad list of principles that the contracting agents decide directly, but also sequentially, from the original position: (i) institutional principles, (ii) the principle of fairness, (iii) principles of individual duty – principle of maleficence, principle of mutual aid, principle of consideration, principle of paternalistic guidance, and principle of establishing a natural duty of justice, and (iv) principles of supererogation.<sup>8</sup>

Basically, Richards believes that when contracting agents are placed in a contracting situation similar to Rawls' (which includes a veil of ignorance), they will all agree upon the moral principles (i)-(iv). It is from these principles that we – i.e., the non-contracting agents – can decide whether or not various actions, both actual and hypothetical, are in conformance with these principles. If an action is not in conformance with any of these principles, then it is morally impermissible. Otherwise, an action is morally permissible.

Just as I do with Rawls' account, I will argue in this dissertation that Richards' principles, as objects of agreement by the contracting agents, will not be sufficiently or adequately applicable to specific moral choice situations. There are various reasons why these principles, as principles, are problematic. Chapter 3 of this dissertation is an extended argument against the idea that the objects of agreement are principles or rules.

---

<sup>8</sup> David Richards. *A Theory of Reasons for Action*. (1971) See pp. 92-95. Note that the principles of individual duty, as well as of supererogation, are decided independently of the existence of institutions, such that these would be agreed upon even if the contractors believed that no institutions (no government) were to exist or ought to exist. Richards discusses this on p. 92.

#### *1.1.4 Non-Rawlsian contractarian ethical theories*

The previous subsection is perhaps ironic. I present my account as Rawlsian, and yet I reject what some may consider central to Rawls' theory – i.e, his theory as a theory of principles of justice. This criticism misses what I think to be a central strength in Rawls' theory. This strength is found within his fundamental contractarian theory, where it is contracting agents who are placed behind a veil of ignorance in order to make normative decisions. Much progress, I believe, can be made in normative and applied ethics with a theory where the contracting agents make decisions while placed behind a veil of ignorance.

Of course not all philosophers think that a Rawlsian veil of ignorance will lead to the conclusions (i.e., principles, in Rawls' case) that Rawls thinks. I am with these philosophers on this point. Yet, I am not with them on what the contracting agents would agree to, on the supposition that they do not agree with Rawls. For example, we have Harsanyi who thinks that while behind a veil of ignorance, the contracting agents would agree to principles which maximized average utility as a principle governing morality. That is, Harsanyi argues that the product of agreement from behind a veil will be a fundamental ethical principle of maximizing average utility. I disagree with Harsanyi on this point. My disagreement will become more apparent in chapters 3 and 4 when I characterize in detail Rawlsian ethical act contractarianism.

Another philosopher proposing contractarianism as an ethical theory is T.M. Scanlon. He rejects the veil of ignorance as a constraint placed on the information possessed by the contracting agents. His view is something like this: X is permissible if and only if it conforms with any given principle which has not been ruled out on the grounds that some agent has rejected it on reasonable grounds. Scanlon's view is admirable and yet mysterious. I will give a criticism of it in the second chapter of this dissertation. I should also note another contractarian theory as an ethical theory. This is Gauthier's theory. Its relevance as a competitor will be illustrated in the second and third chapters, and I will leave it at these places to characterize it.

## 1.2 RAWLSIAN ETHICAL ACT CONTRACTARIANISM

I've now explained the issue/question which this dissertation is addressing. Since the kind of ethical theory I'm defending is contractarian, I've given background on contractarianism in general. Furthermore, I've given a brief characterization of the popular contractarian ethical theories in the literature. Now, in contrast to these theories, I propose (and will defend the following alternative):

Rawlsian ethical act contractarianism: An action A is morally permissible if and only if, from behind a veil of ignorance, A would be agreed to by all contracting agents, each of which are the way they actually are in real life, except that they are perfect expected individual utility maximizers.

I will offer two reminders, as well as one further point.

The first reminder concerns the contrast between act versus rule contractarianism. Whereas for Rawls and Richards, as well as for Scanlon and Gauthier, the objects of agreement for the contracting agents are principles/rules, according to the above proposed account, the objects of agreement are not principles/rules but instead actions as agreed upon with respect to specific moral choice situations. Understandably, the idea that "actions are agreed upon" by the contracting agents is a bit mysterious in contrast to principle/rules being agreed to. But I will spell out what is meant by the idea of "actions being agreed upon" in chapter 3.

The second reminder concerns the veil of ignorance. Unlike Rawls and Richards, theorists such as Scanlon and Gauthier oppose the veil of ignorance as an unwarranted constraint on the information provided to the contracting agents as they make their decisions. Recall that by being placed behind a veil, the contracting agents are unaware of their identity in specific moral choice situations. For example, if one master and one slave were to decide whether slavery is permissible, and if they were placed behind a veil, then neither would know whether they were slave or a master, but would know that one of them is a slave while the other master. As evident

from my characterization of Rawlsian ethical act contractarianism, I disagree with Scanlon and Gauthier; the veil of ignorance is essential to a contractarian ethical theory.

Now for the point that I promised to make. Notice from the above characterization of Rawlsian ethical act contractarianism that I have also departed from Rawls in terms of how he characterizes/idealizes the contracting agents. His contracting agents are rational, self-interested, and mutually unconcerned. My contracting agents are just as they are in real life, except for being perfect expected individual utility maximizers. In real life, there are plenty of people who care about others and care that others do well. For example, just think of the many people who prioritize their familial roles as parents or their roles as friends. Utilities can be assigned to outcomes based on these facts. We can easily suppose that some person, say Seth, cares very much that another person, say Rebekah, does well. Suppose that Mark may do something which could make Rebekah worse off. Then such an action on Mark's part would also make things worse off for Seth, all other things equal. And all other things equal, the outcome of Mark's action for Seth, as well as for Rebekah, will have negative utility. In chapter 3 and chapter 4, in particular, I will motivate and argue for the characterization of the contracting as perfect expected individual utility maximizers.

### 1.3 METHODOLOGY AND SKETCH OF THE DISSERTATION

The methodology of this dissertation is fairly straightforward. I argue for my account by considering alternative accounts, and then explaining how they are problematic. I then show how my account is not problematic with respect to those same dimensions. I also consider objections and possible problems to various features of my account, thereby explaining why these objections and problems are spurious. Furthermore, I also consider moral cases in which my account gets results (concerning permissibility and impermissibility) which align with our shared intuitions.

The previous point concerning intuitions is worthy of expansion. First, there may be some who deny the strength of appealing to intuitions in arguing for theoretical claims, and especially in ethics. To this, I can only say that the evaluation of theoretical claims in ethics has

to start somewhere. Second, I think that my appeal to intuition is rather limited. Note that the appeal is to *shared* intuition. Most of us share the intuition that it is impermissible for a surgeon to kidnap and kill an innocent person for the use of transplantation of organs to save the lives of five innocent dying people. It is a mark against a theory whose results do not align with this shared intuition. Note also that I will do my best to appeal to shared intuitions in *moral cases*. There are other things that we have shared intuitions about than just cases. For example, we have intuitions about explanatory adequacy. Consider a simple version of ethical egoism, and the action of someone, A, torturing a dog for the mere fun of it. According to a simple version of ethical egoism, if A's action is in any way wrong, it is wrong because in torturing the dog, A is not best promoting her self-interest. For many of us, we share the intuition that this is not the correct explanation for why A's action is wrong. Now, in this case, I think that this appeal to shared intuition can be used to evaluate an ethical theory. But not all appeals to shared intuitions about explanatory adequacy may be good indicators of the goodness of a theory or theoretical claim. Thus, I will do my best to limit the appeals that I make to shared intuitions about things other than moral cases.

Now I will give a sketch of the dissertation. Since my Rawlsian account of course contrasts with other contractarian accounts, I will consider other contractarian (ethical) accounts. The two primary candidates that I explore in order to give credence to my own are Gauthier's Hobbesian-style ethical theory and Scanlon's Kantian-inspired theory. In **chapter 2**, I explain the deficiencies of each account with respect to certain problems typically endemic to contractarianism. One such problem is called the Outlier Problem. The problem is as follows: Contractarianism assumes that the moral contract is made by parties to the contract *agreeing* to the contract in light of what is *mutually beneficial*. Now, there are those, such as non-human animals and persons with severe mental disabilities, who are incapable of making agreements. And there are others who, although can make agreements, have less to offer than what they need in return; this includes some persons with severe physical disabilities. Both groups are outliers to

the contract, as they are excluded either on the grounds of being incapable of making agreements or by being incapable of (mutually) benefitting others. I argue that a Hobbesian account, such as Gauthier's, cannot give an adequate solution to the Outlier Problem. My account, on the other hand, successfully solves the problem, as it employs a veil of ignorance. From behind a veil of ignorance, one does not know whether she is a trustee for non-human animals and those with severe mental disabilities, and as such will make decisions in light of the possibility of being a trustee. And by being placed behind a veil of ignorance, one does not know whether she is a person with a severe physical disability. According to my account, there are no outliers.

In **chapter 3**, I argue for what I take to be the largest departure from not just Rawls' theory, but really any existing contractarian account. Namely, I argue that for an act-based theory opposed to a rule-based one. As I've explained, contractarianism has typically been formulated as a view in which contracting agents agree upon rules or principles. And it is with these rules that actions are then evaluated as morally permissible or impermissible. In contrast to the typical rule-based view, I contend that the contracting agents decide directly on actions. Of course, I will have to explain what is meant by "agreeing to an action." I also explain more specifically how an act contractarian view works. My argument for the act-based component of the theory is inspired by the rule versus act utilitarianism debate. I end up arguing that rule contractarianism either is not a genuine alternative view or if it is a genuine alternative, it either leads to counterintuitive results or is guilty of rule-worship.

Because the central feature in which I agree with Rawls is his endorsement of a veil of ignorance, and since there have been various objections to contractarian accounts employing a veil of ignorance, I will have to consider and reply to these objections. It is in **chapter 4** that I address these objections, arguing that none of them pose a serious problem for my account. Some objections to the veil of ignorance have relied on an assumption that by reasoning behind a veil of ignorance, contracting agents would make decisions according to maximin, or would arrive at the Difference Principle. These principles have had counterexamples leveled against them. But as I

point out, the contracting agents need not reason according to maximin (which leads to the Difference Principle) while behind a veil of ignorance. I show that since the contracting agents of Rawlsian ethical act contractarianism reason according to the principle of maximizing expected individual utility, and since they consider sophisticated utility assignments, they are able to arrive at decisions which align with our shared intuitions. Other objectors to the veil of ignorance argue that certain aspects of our individuality are not respected when morality is determined by contracting agents from behind a veil. But once again, by making sophisticated utility assignments, contracting agents are able to make decisions which do, in fact, respect our individuality. Finally, in **chapter 5**, I discuss the conclusions of the dissertation, as well as point to various areas of research relevant to making Rawlsian ethical act contractarianism a more plausible view.

## CHAPTER 2 – NON-RAWLSIAN ETHICAL CONTRACTARIANISM

### 2.1 INTRODUCTION

In the previous chapter, I introduced the basic idea of the contractarian account that I'm defending in this dissertation. This contractarian account is intended to be an ethical theory, as opposed to a political theory, or theory of distributive justice. Although I depart from Rawls with respect to this last point, my account is Rawlsian in that it employs a veil of ignorance. In this chapter, I consider two other ethical contractarian accounts, ones which do not employ a veil of ignorance. The first is Hobbesian in character, and the particular form of it that I consider is one offered by Gauthier. The second is Kantian in character, and the particular form of it that I consider is one offered by Scanlon.<sup>9</sup> I will discuss how both accounts are problematic.

The evaluation of these two important accounts of contractarianism is done in light of how well they handle three problems that I introduce. I introduce these problems throughout the discussion of these accounts. In particular, I introduce what I call the *gangster-inclusion problem*, as well as the more well-known *outlier problem* while discussing Gauthier's account. Although I also discuss Scanlon's theory in light of the latter two problems, I ultimately focus on his account in light of a third problem, which I call the *numbers problem*. Although I do not discuss what I find to be the solution to these problems until chapter 4, I think it is important that I note to the reader that ultimately I think that adequate solutions to each of these three problems together, simultaneously, requires introducing a veil of ignorance. The two main kinds of contractarian accounts that I consider in this chapter do not endorse/employ a veil of ignorance.

### 2.2 GAUTHIER'S HOBBSIAN THEORY

#### 2.2.1 *Characterization of Gauthier's Hobbesian theory*

---

<sup>9</sup> Note that some would characterize Rawls' theory as essentially Kantian, and thus by my endorsement, I have condemned my own view. This is not correct. Even if Rawls' theory is Kantian, the features that I employ from his theory will not suffer from the same problems that I believe are endemic to Scanlon's theory.

In this subsection, I characterize Gauthier's Hobbesian theory, which for now on I will just call Gauthier's contractarianism.

Gauthier's contractarianism: An action is morally permissible if and only if it is not forbidden by any principle which the contracting agents would agree to, where such contracting agents are rational, mutually unconcerned perfect individual utility maximizers, and have knowledge about their own personal capacities and preferences (more specifically, they are not placed behind a veil of ignorance).<sup>10</sup>

A few notes must be made. First of all, the nature of the contract is hypothetical according to Gauthier. Although Gauthier does not rule out the moral relevance of actual agreements made between people, whether or not actual agreement is morally binding is determined by what the contracting agents would agree to – i.e., the moral relevance of actual agreement is based on the hypothetical agreement of the contracting agents. A second note to make concerns the contracting agents. These agents are human beings, although Gauthier doesn't say anything explicitly that would rule out the possibility of making contracts with intelligent alien individuals. A third note is this. The contracting agents are not concerned with maximizing overall societal utility, unless in doing so each is convinced that she will fare best under a principle of maximizing societal total. (This is highly unlikely, though.) They are concerned with maximizing their own utility, not with maximizing the utility of others. Gauthier believes that the contracting agents will agree to constraints on certain activities in the pursuit of maximizing (individual) utility. This will be rational because such constraints encourage mutually beneficial enterprises that would not exist without such constraints. Gauthier identifies these constraints with the content of morality; they are rational constraints, and as moral constraints they are a product of rational agreement. Also, Gauthier's moral constraints are not necessarily identical to the constraints endorsed by commonsense morality, although there will be some similarities. Any

---

<sup>10</sup> This characterization is gathered from various pages and passages from David Gauthier's *Morals By Agreement*. Clarendon Press; Oxford University Press: 1986.

divergence from commonsense morality then means that Gauthier's theory is revisionist, which he is fine with.

### *2.2.2 The gangster-inclusion problem and structure of the argument*

In this subsection, I explain what I take to be a problem for contractarian accounts like Gauthier's. Let me at this point first give a rough statement of the problem. The objection has to do with how equal or unequal the contracting agents are amongst each other. If it were the case that the contracting agents were equal (or sufficiently equal), then as it will turn out, there would be no problem. They would not be able to level credible threats against each other in order to get the threatened parties to agree to what they desire (agreements which may run counter to our moral intuitions). I think, though, that there is good reason to think that the contracting agents are not equal, and that furthermore, there can plausibly be said to be enough disparity in strength between some and others among the group of contracting agents that credible threats could be made by the stronger upon the weaker such that counterintuitive agreements would indeed result.

In order to make sense of this problem, I should first explain what position one may take on the differences that hold between contracting agents, and, in particular, whether they are relevantly, approximately equal to each other.<sup>11</sup> If the contracting agents are equal, then there should be no problem with respect to getting them to come to agreement, at least respect to the problem that I have in mind. The problem, once again, is one where those with great strength over others will level credible threats in order to get their way. But if everyone is equal, then this possibility is eliminated, and credible threats are eliminated. Thus, if a contractarian account can maintain plausibly that such equality can exist between contracting agents (without a veil), then it will be able to avoid the problem that I have in mind – that is, agreements being made in light of credible threats. And so, it is relevant that I discuss the possibility of equality amongst contracting agents (without a veil).

---

<sup>11</sup> It is important that I do not mean by "equality" here "moral equality." Instead, the kind of equality that I have in mind is non-moral equality. It is possible that two moral agents are unequal in the non-moral sense and yet, of course, equal in the moral sense.

I should be more specific about the kind of equality that I have in mind which would avoid the problem of agreement in light of credible threats. The kind of equality I have in mind is best characterized as follows:

Relevant, approximate equality: All contracting agents are equal in the sense that each are able to comprehend all the facts surrounding the proposed contract and are able to protect themselves against force in an approximately equal way. Even if there are some people that cannot initially defend themselves from harm in every encounter, these people are able to retaliate in a way that is a reason of deterrence for others to aggress. Relevant, approximate equality is the case between contracting agents just so long as no one or group of persons has a bargaining advantage over any others. Furthermore, wherever there is one-to-one inequality, there is always a group of people who would band together to suppress this inequality.

If we consult our own experience, we know that there are people who are not hesitant about the use of force to get their way. In real life, these people certainly do exist. It's not that their existence is necessary. But if we consult history, as well as present sociological studies, we see that their existence is actual. I think that any contractarian theory must be constructed in light of these facts, and not by dismissing these facts by assuming something otherwise.

These people, who are not hesitant about using force to get their way, I will call "gangsters." But this is, of course, ambiguous. Depending on what's at stake, there are many people who are willing to use force to get what they want. What is furthermore essential to a gangster is their ability to get what they want upon using force, or by threats of force. So, a second property of gangsters is this: Their force, whether individual or collective, is such that they are able to (i) successfully win in many given situations against other individuals and collections of individuals (and of great size) and (ii) be able to successfully win in many given acts of retaliation, even by collections of individuals (and of great size). It is the gangsters, as contracting agents, who hold a bargaining advantage in the contracting process. Note that if

gangsters exist, then the characterization of all people being relevantly, approximately equal is false. And as such, there is no basis, as far as I know, (with the exception that the contracting agents are placed behind a veil of ignorance) for characterizing the contracting agents as being relevantly, approximately equal.

So, the point that I'm making is this: Given that gangsters are present in real-life, and given that contracting agents are to be composed of real-life agents, it appears that such gangsters will be part of the contracting process. Now, if one were to deny this by upholding the idea that there is relevant, approximate equality between contracting agents, then, of course, they would bypass the problem of credible threats being introduced in the contracting process in order to come to results (which would be counterintuitive). But this route is implausible. And Gauthier agrees; he is a realist about the variety of contracting agents, which includes the presence of what I've called "gangsters." Given this, Gauthier is challenged by what I will call the gangster-inclusion problem. I will now characterize the problem as follows:

The gangster-inclusion problem: Given that some contracting agents are gangsters, the results of the contract can be determined by credible threats made by the gangsters. But inclusion of the gangsters, with their credible threats, can result in a morality that very much diverges from our intuitions.

I will now propose the following structure of the argument against Gauthier's account in light of the gangster-inclusion problem:

Structure of the argument against Gauthier's account in light of the gangster-inclusion problem

- (1) Either one should maintain relevant, approximate equality between the contracting agents or instead maintain that there is relevantly strong inequality.
- (2) One should not maintain relevant, approximate equality between the contracting agents.
- (3) Therefore, one should maintain relevantly strong inequality between the contracting agents.

- (4) If one maintains that the contracting agents have full knowledge of their capacities and preferences and if (3) is true, then one's contractarian account will not be able to provide a solution to the gangster-inclusion problem.
- (5) Gauthier maintains that the contracting agents have full knowledge of their capacities and preferences.
- (6) Given that (3) is true (and, in fact, Gauthier believes it to be true) and (5) is true, Gauthier's contractarian account will not be able to provide a solution to the gangster-inclusion problem.

We have already discussed (1)-(3) in some way. In addition, it should be apparent that (5) is true; Gauthier does not endorse a veil of ignorance, believing that the contracting agents should have full knowledge of their own capacities and preferences. This means that each contracting agent will know of their strengths and weaknesses in relation to other contracting agents. And since (6) is the conclusion, the only remaining step to consider is (4). I will in the next subsection show that (4) is true, despite Gauthier maintaining seriously both the inequality of the contracting agents and their knowledge of their particular characteristics constituting this inequality.

### *2.2.3 Gauthier with respect to the gangster-inclusion problem*

I have just stated that if there is relevantly strong inequality between contracting agents (say, with the presence of gangsters) and if the contracting agents have full knowledge of their preferences and capacities (for example, their strengths), then indeed there is the gangster-inclusion problem, where the gangsters can level credible threats in order to get their way in terms of what's agreed upon. In this section, I will argue that Gauthier cannot bypass this problem.

The best way to directly respond to Gauthier is by criticizing the reasoning of his contracting agents which is supposed to avoid counterintuitive results. Gauthier thinks that despite the inequality and the full knowledge of contracting agents of their strengths and weaknesses, there will nevertheless be an agreement in assessing outcomes in light of the

Lockean proviso constraining the behavior of individuals (and as such, it is used as a baseline for comparing contracts agreed upon). Gauthier has this to say in characterizing the Lockean proviso:

We interpret the Lockean proviso so that it prohibits worsening the situation of another person, except to avoid worsening one's own through interaction with that person. Or, we may conveniently say, the proviso prohibits bettering one's situation through interaction that worsens the situation of another. This, we claim, expresses the underlying idea of not taking advantage.<sup>12</sup>

Now, the contracting agents will reason in such a way that employs something like a Lockean proviso. According to Gauthier, the stronger of the contracting agents will reason in such a way that it will not violate the Lockean proviso in terms of themselves, but in doing so, this will mean that the weaker will also agree to terms which will mean that they are not violated according to the Lockean proviso. The idea is this: Neither the stronger nor the weaker will be made worse off according to the Lockean proviso according to the agreements they make, despite the disparity in strengths/capacities between the contracting agents.

Of course, the Lockean proviso as a rational constraint on bargaining has to be motivated. Gauthier motivates it by having us imagine a society in which there are masters and slaves, a sort of state of nature in which there, indeed, is considerable inequality. (Once again, Gauthier and I are in agreement concerning inequality). The masters and slaves are as follows: The masters are supposedly tired of being masters, and really just want to achieve the benefits of being masters, not having to put up with the costs of being masters. More obviously, the slaves are tiresome of being slaves, and they desire a better life than that of being slaves. For both parties, the present arrangement is sub-optimal.<sup>13</sup>

Suppose that the masters understand the sub-optimality of their situation. And as such they propose to the slaves their freedom on the condition that they are to be servants of some

---

<sup>12</sup> Gauthier, p. 205.

<sup>13</sup> Ibid., pp. 190-192.

kind. Upon agreement, the servants would enjoy some rights, as well as some pay for their services.

But then, Gauthier argues that were the slaves to agree to this, they would not comply with the agreement, since the terms were agreed to from a position of coercion in the first place. The newly made servants would say to themselves – Look, we are not going to comply with an agreement that was made under conditions of coercion. So, we should rebel against the terms of agreement. And more importantly, the slaves would rebel to the point that they would be able to take control, at least partially, over the masters. As such, the masters, knowing that the slaves would respond in such a way, would not make the original proposal in the first place. Thus, the baseline for comparing contracts is not one where the conditions are that of coercion by some over others. Instead, the baseline should be one where coercion is absent. If coercive conditions define the baseline, then upon agreement the former slaves would not comply. So, once again the baseline for comparing contracts is one characterized by the lack of coercion. And this, according to Gauthier, is a baseline characterized by the Lockean proviso. (A state in which persons abide by the Lockean proviso is a state absent of coercion.

Now, before giving my criticism of Gauthier's account it is important that I substitute talk of "masters" with "gangsters." I think that, first of all, this substitution reflects present-day concerns. It also, more specifically, reflects how supposedly free individuals in a society (unlike slaves) may have to succumb to the demands of others, namely "gangsters."

I am now ready to give criticism of Gauthier's account. The way that I show this is by showing that the situation characterizing the masters is not sub-optimal. Gauthier's argument goes through only if the situation of the masters/gangsters is sub-optimal. But as I'll argue, their situation is optimal for them. Given that Gauthier agrees that there will be inequality between contracting agents, and given that they will have knowledge of their own capacities, and given that the situation of the stronger (the masters/gangsters) is optimal, Gauthier's account is not able to solve the gangster-inclusion; in fact, it suffers from it.

Establishing optimality of the gangsters/masters. First of all, it assumed that, in my words, “The masters have devised a coercive apparatus over their slaves that is costly to maintain.” But this is just not the case. Although there are costs, the net evaluation is such that the masters gain so much more from maintaining this coercion than otherwise. Recall the argument about rational agreement based on rational compliance. The slaves, upon becoming servants, will not comply with their servitude. The result is this: The servants would be able to destroy some of the property of their former masters, as well as accumulate property from their former masters in ways that are more beneficial to the slaves than servitude.

Secondly, it is assumed that, in my words, “slaves do not work for their masters at an optimal level.” Making this judgment requires that the masters know what would be optimal, or at least more optimal. Given this knowledge, they would be able to level threats against their slaves in order to increase optimality, if not perhaps to maximize it. In response, some might say that production is maximized when workers don’t feel coerced. I think that this might be true with respect to work that involves interaction with customers, in particular sales. In order to maximally make sales with customers, the salesperson might have to feel at ease, or minimally comfortable. Yet this might be countered with the idea that salespersons often work under the pressure of being paid solely, or mostly, via commissions on sales. What, then, if the commission for sales was being able to keep one’s life? As for other work, we should consider hard labor. Here, the masters would seem to be able to maximize production by “whipping out” of their slaves maximal production.

A third reason for thinking that the position that the masters (or: gangsters) are in a position that is not sub-optimal, but in fact perfectly optimal for them, is this: It was assumed that it was a cost to the masters to use violent measures on their slaves. But why think this? To many of us, this would indeed be costly. But we are not the paradigms of masters. These masters, I say, are not at all against using violent measures to reinforce obedience. (Here, think of the gangsters that I have characterized.) In fact, perhaps some enjoy it. I know some people

like this. Nevertheless, we might suppose that since the contracting agents are mutually unconcerned, they do not enjoy inflicting punishment and enforcing threats on their slaves. In that case, we should think about how the masters (and/or gangsters) go about their business. Instead of having to witness punishment, they can just hire others to do their dirty work for them. Perhaps they could even get some of the slaves (let's call them "trustees") to do their punishment work for them. As such, it is not costly to the masters to maintain their slaves' obedience; they just hire other people to maintain obedience.

I've now argued that inequality between the contracting agents will be such that reasoning by them according to the Lockean proviso will not lead to results which align with our shared moral intuitions. The stronger – the masters or the gangsters – will not reason in the way that Gauthier describes. The masters and/or gangsters are not motivated to make agreements aligning with our shared intuitions. It will be in the interest of both parties to stay in slavery (if we are talking about masters) or constant exploitation (if we are talking about gangsters).<sup>14</sup>

Conclusion: Gauthier's contractarianism does not solve the gangster-inclusion problem. The gangsters perhaps do not suffer uncompensated costs by enslaving others. Second, it is the gangsters and not the slaves who have the overriding power to not comply with the contract – i.e., it is in the interest of the gangsters to not comply with the contract that they themselves have proposed. So, if the slaves were to propose this contract, the gangsters would not agree, laughing.

---

<sup>14</sup> There is one final thing to consider. Suppose that I'm wrong about the optimality of the masters being in scenario 1. Or even more poignant, suppose that the masters for some reason mistakenly think that scenario 1 is not the best for them (or: should not be used as a baseline for comparison). Then now let's suppose that they dismantle their coercive apparatus under agreement with what they might think is more optimal for themselves. Then, on the supposition that they then realize that the scenario they find themselves in is sub-optimal, I think this: I think that it is quite plausible to think that the gangsters (or: masters) are so powerful and cunning that they are able to recover from almost anything (except the death of all of them). Suppose that the gangsters/masters dismantle all of their coercive apparatus upon making the slave-to-servant agreement. Does this entail their inability to recollect their powers to coerce their former slaves? As a logical point, no. Furthermore, and more importantly, it just seems to be the case that there are certain groups of people (namely, those who I have characterized as "gangsters") who can organize their powers in such a way that they always "win." The idea is this: If the slaves-to-servants do not comply, then the gangsters/masters will then reinitiate their coercive abilities in such a way that the slaves-to-servants lose. Now, this is common knowledge, and so, it is rational for the slaves to agree to whatever contract proposed to them, just so long as it doesn't make themselves worse off than where they are in the state of nature.

Third, the gangsters/masters are able to respond beneficially to the non-compliance of their slaves: If the slaves don't comply, then they (the gangsters/masters) will just re-employ their coercive apparatus against the slaves, and they will do it successfully.

#### *2.2.4 The outlier problem*

In this subsection, I introduce what's been called the outlier problem for contractarian theories in general. Since the outlier problem concerns how various ethical theories get the moral status of certain creatures (beings) wrong, I will first discuss the notion of moral status. We can characterize every being, or object, as having a moral status. If it doesn't matter how we treat a being – i.e., it is not possible for us to wrong it in any way – then that being's moral status is that it has no moral standing. For example, the moral status of a rock is that it has no moral standing. If I throw a rock and smash it against the concrete, I may have done something wrong, but not in virtue of wronging the rock or the concrete. Perhaps the concrete was part of my neighbor's driveway, or the rock I smashed was from my neighbor's rock garden. I've wronged my neighbor, not the rocks or the concrete. In contrast to rocks, beings that have the highest moral status are persons.<sup>15</sup> Although it is controversial what exactly counts as a standard for personhood, there are some fairly fixed intuitions concerning examples of persons. For example, it's intuitive that human beings with at least some minimal self-awareness are persons. It's also intuitive that if we were to come into contact with self-aware aliens, then we ought to consider them to be persons. It's more controversial whether non-human animals, who are not self-aware, are persons. But that they have some moral standing should not be controversial. Non-human animals having a central nervous system are certainly sentient, and as such, these beings have at least some moral standing. This is also the case for those human beings with such severe mental disabilities that they are not self-aware; although they lack self-awareness, they are certainly aware of pain and pleasure, and so they have some moral standing.

---

<sup>15</sup> Of course, it is possible that there are beings whose moral status is even greater than persons. These beings might be so great in moral status that we might be required by morality to become their servants. But thus far, there is no good evidence indicating that these beings are actual.

The examples that I've just listed concerning beings with various moral statuses are pre-theoretical commitments for any ethical theory. An ethical theory should not have the result, nor make the assumption, that it is not possible to wrong a sentient non-human animal. And it should not get the result that it's possible to wrong a rock. But this is not all. As we will see, the outlier problem also concerns human beings who *are* self-aware, and as such are uncontroversially persons, despite prejudices and misunderstandings of the past.<sup>16</sup>

In total, I have in mind four different groups of beings with which the outlier problem is concerned (or: should be concerned); three of these groups of beings are almost certainly persons, and the fourth of these groups certainly have moral standing, although it is still controversial whether they are persons.

Without making any pretense to sophisticated training in psychology or medicine, I will define the first two of the four groups of beings as follows:

Sufficiently Severely Physically Disabled (SSPD): A person P is sufficiently severely physically disabled if and only if the cost of providing services for P to have a minimally decent life is greater than any possible total benefits that can be gotten from P.

This is of course not very precise, but an attempt at greater precision would require employing notions from professional psychology. But such precision I don't think is needed for purposes of making clear the very real outlier problem. Instead, let's just get concrete about who would count as an SSPD. Although Stephen Hawking would be someone we might think of as sufficiently severely physically disabled, he, in fact, is not. This is because the benefits that we receive from him in his work on theoretical physics outweigh the benefits he must receive. What's essential to being a SSPD is that one not be able to reciprocate in a way in terms of goods and services that we think typical between "ordinary people." Perhaps an example of someone who is sufficiently

---

<sup>16</sup> The reason that I say "despite prejudices and misunderstandings of the past" is this: I suspect that some people, or perhaps a majority of people, have considered other people, fully self-aware, not to be persons. We can imagine that some severely retarded people and some physically disfigured people (who were otherwise "perfectly" intelligent) being treated as something like animals.

severely physically disabled would be someone just like Stephen Hawking, except that person had only average intellectual capacities.<sup>17</sup> I don't want this example to be misleading as a characterization of all SSPDs. Not all SSPDs are wheel-chair bound, and not all wheel-chair bound people are SSPDs (as I characterize the concept). We could imagine someone who is able to work 8 hours a day, but then requires both expensive medication and constant care for the other 16 hours such that it costs more to provide these services than the benefits gotten from her by her work.

The second group of beings with which the outlier problems concerns is as follows:

Sufficiently Low Rational Capacities (SLRC): A person P has sufficiently low rational capacities if and only if P is not capable of making rational decisions which significantly affects P.

This definition is unfortunately too imprecise. So I will say some things which should avoid confusion. To get an idea of those who would have sufficiently low rational capacities would be those with an IQ under 70, or those who have court-appointed legal guardians; of course there are counterexamples, but that is not the point. As an example, I have in mind someone with sufficiently low rational capacities who is indeed capable of deciding that she wants a bicycle. But she may not be capable of devising a plan on her own on how to acquire one. She may need assistance in understanding that she needs a part-time job to earn money and save for one. More importantly, in the absence of worker rights laws, she will need someone to advocate for her in searching for a job. Without such assistance, and in a society with no workers rights laws, she may just agree to a ditch-digging job for \$2/hr. Making such an agreement would be irrational on her part.

There is a third group of persons that I'd like to mention, relevant to the outlier problem.

This third group is children and infants. Infants could perhaps be classified as SLRCs, but they

---

<sup>17</sup> I want to be careful here in my use of "average intellectual capacities." I'm aware that there are very many different kinds of intelligence. But with each kind of intelligence, we can still make (at least roughly) sense of the idea of degree.

seem to differ from SLRCs in this fashion: Some might think that SLRCs are full-fledged persons, and yet think that infants are not full-fledged, even if perhaps they have some moral standing. Children, I consider, are those who are somewhere between the age of infancy and adolescence. Children, too, could be considered to be SLRCs, as they are incapable of making rational decisions which affect them in some of the most significant ways. But most children become, say by the age of at least 16, not SLRCs, but ordinary people. In effect, both infants and children are incapable of making rational decisions that is required of those who are party to the contract. For this reason, I will distinguish SSPDs and SLRCs from “children” (to include children under the age of 12 and infants). I want to make it clear that I am not distinguishing children (as opposed to infants) from the category of SLRCs for the reason that they are not persons fully, whereas SLRCs are. This is just not the case, in my opinion. Children (as opposed to infants) are fully persons. The reason for distinguishing them from SLRCs is that they, and not SLRCs, develop into ordinary people. Infants, who I have included under the category of “children” broadly construed, perhaps are not persons, or at least not fully persons like SLRCs. Although they differ from children in this way, they are similar to children in this way: Just as most children will develop into ordinary people, so too will infants.<sup>18</sup>

And finally there is the fourth group of those who have moral standing that must be mentioned. These are non-human animals. Once again, they could be included within SLRCs. For example, if they are considered to be persons just like any other human being, then it would be correct to consider them as SLRCs. But not everyone considers non-human animals to have the same moral standing as human beings. As such, it would be correct to consider them as a separate category; although, I think it is correct to at least minimally consider them to have moral

---

<sup>18</sup> Note that what I say here is consistent with a position that says that the outlier problem does not concern fetuses who are not sentient and/or viable. This is because one can reasonably hold that non-sentient and/or non-viable fetuses have no moral standing whatsoever. Infants, on the other hand, have at least some moral standing, or at least this is what is typically thought.

standing. As such, we have now four categories to consider: SSPDs, SLRCs, children, and non-human animals.

The outlier problem is as follows. Assuming that the moral contract is made by parties to the contract *agreeing* to the contract in light of what is *mutually beneficial*, certain beings which we think have moral standing will either (i) be excluded from the contract or (ii) upon being included in the contract, will be disadvantaged in the contracting process. Although with (i) it is more obvious why such parties would be called “outliers”, it is not as obvious why those in category (ii) would be called as such. Rhetorically put, if someone is part of the contracting process, then how could they be an outlier to the contract? The answer is that although they are part of the contract, what they would agree to given their disadvantage would be tantamount to being treated as if they were an outlier to the contracting process. Although SSPDs can understand the terms of the contract just as ordinary persons can, they are disadvantaged in the sense that they have less to offer others than what they need in return.

Returning to the characterization of the outlier problem: While some beings having moral standing are not capable of making rational agreements, others are not capable of benefiting others in a way that matches the benefits that they need to receive. Examples of those in the first case are non-human animals, persons with severe mental disabilities (SLRCs), and children. Those in the second case are the sufficiently severely physically disabled. At least some of these people are “perfectly” rational. The problem is not their ability to make agreements or even to make rational agreements. Instead, it is that there is no (purely non-moral) reason for making contracts with them. Or, as discussed above, even if there is a reason to make contracts with them, this will not be for reasons of mutual benefit. Instead, one would enter into a contract with them in order to get them to agree to terms that don’t benefit them (i.e., SSPDs). If, by definition, the benefits received from them cannot match the benefits they need to receive,

then why include them in the contracting making process?<sup>19</sup> Or, if upon including them into the contracting process, why agree to terms that would allow them to get benefits that outmatch the benefits that could possibly be gotten from them? The outlier problem amounts to somehow including into morality the outliers to the contract; or, in the case of SSPDs, it amounts to showing that they will get in return more than would be allowed by the bargaining disadvantage against them. Presented as an objection, the outlier problem is an exclusion objection, in that contractarianism excludes certain beings from the contracting process in such a way that they have not been recognized as having the appropriate level of moral standing..

### 2.2.5 *Gauthier (and Morris) with respect to the outlier problem*

In this subsection I show that Gauthier's contractarianism faces worries regarding the outlier problem. The defender of his theory that I will explore and evaluate is Christopher Morris.

Morris explains that:

....some moral theories – namely, certain contractarian theories – do not regard the possession of any natural or independent properties as sufficient for moral standing. For such theories, moral standing requires in addition that individuals be related to one another in certain ways.<sup>20</sup>

As such, Morris reminds us to distinguish between two main ways a being may have moral standing. The first, and more traditional, way by which a being has moral standing is by it possessing some *non-relational* property – e.g., rationality, sentience, being a member of some species, having a soul, etc. The second way a being possesses moral standing is by it possessing some *relational* property.<sup>21</sup> What matters here is that a being is related in a certain way to other

---

<sup>19</sup> I must note, importantly, here that not all contractarian accounts necessarily suffer from this problem. Some contractarian accounts stipulate from the get-go that mutual benefit is not necessary in order to be included in the contracting process, as well as receiving protection and benefit from the contract itself (i.e., by everyone following through what is required by the contract). As we will see later, T.M. Scanlon proposes a contractarian account that does not require mutual benefit as a condition for what's agreed upon as the moral contract.

<sup>20</sup>Christopher Morris. "Moral Standing and Rational-Choice Contractarianism." In ed. Peter Vallentyne. *Contractarianism and Rational Choice*. Cambridge University Press: 1991, p. 77.

<sup>21</sup> Morris, p. 77.

beings. Morris endorses this latter way of granting moral standing. More specifically, he develops a Gauthier-inspired notion of moral standing as follows:

*Primary moral standing:* A has moral standing in relation to B if (1) A and B are in the circumstances of justice, (2) A and B are capable of imposing constraints on their behavior toward one another, and (3) A so constrains his or her behavior toward B.

*Secondary moral standing:* A has moral standing in relation to B if (1) B and C are in the circumstances of justice, (2) B and C are capable of constrained action, (3) C constrains his or her acts toward B, and (4) is the object of C's preferences, that is, C cares sufficiently about A that it would not be rational for C to cooperate with B unless A were accorded moral standing in his or her relations with B.<sup>22</sup>

The idea here is that non-human animals, SSPDs, SLRCs, and children are each granted secondary moral standing in virtue of being objects of preference by those with primary moral standing. Now, Morris is clear to remind us that secondary moral standing is not the same thing as having indirect moral standing. Morris reminds us that:

Secondary moral standing, it should be emphasized is merely a manner in which moral standing can be acquired. Someone who acquires moral standing in this way has genuine moral standing. It is secondary only in the sense that were no one to have primary moral standing, no one could have secondary moral standing.<sup>23</sup>

The idea with Morris's account is that genuine moral standing is accorded to the outliers of the contract. More crucially, SSPDs, SLRCs, non-human animals, and children are the objects of preference (as noted by Morris' condition (4)) of some other being that has primary moral standing. As such, they are to be considered in terms of what is agreed upon by the contracting agents (i.e., those with primary moral standing), and will be considered equally to those with primary moral standing; as such, they have genuine moral standing.

---

<sup>22</sup> Ibid., pp. 89-90. This is word-for-word from Morris.

<sup>23</sup> Morris, p. 90.

I think that Morris's Gauthierian solution to the outlier problem is unsatisfactory. In addition to what was said in the previous paragraph, that there is nothing about contractarianism which requires a relational property account of moral standing, I think that there are two problems. With respect to the second problem, I will explicitly state the advantage of a non-relational account of moral standing.

In setting up the first problem, recall that according to Morris's solution SSPDs, SLRCs, children, and non-human animal would be objects of preference by at least some contracting agents. And those who have these preferences would refuse to cooperate (according to Morris) with others who did not constrain themselves in their treatment of the objects of preference – i.e., SSPDs, SLRCs, children, and non-human animals. In the case of SSPDs, who may actually be party to the contract (depending on how to interpret Gauthier and Morris), we could imagine that their disadvantage in the contracting process (where their disadvantage is that they have less to offer than what they need in return) is eliminated by the presence of other contracting agents who have as objects of preference such SSPDs.

Now for the first problem. Let's suppose that every SSPD, SLRC, child, and non-human animals possesses the relevant relational property to at least one person with primary moral standing. If this were the case, then this would be a fact about the world. But this fact is contingent, and so the fact that any given SSPD, SLRC, child or non-human animal has moral standing is also contingent. Of course, such contingency is not problematic if we are talking about a person having moral standing contingent on the fact of their existence, which is contingent itself. (That I exist is a contingent fact.) What's problematic more specifically is that my moral standing would be contingent on whether I am an object of preference of another person. And this, I think, has disastrous results in terms of explaining our intuitions concerning wronging others. To illustrate, suppose that an SLRC is abandoned by her parents, has no siblings, and forms no really close friendships. Then she does not have the kind of moral standing Morris has in mind. Suppose then that some person breaks into her apartment, and takes

from her, her very few possessions, and purposefully destroys her cherished porcelain figurine. Has this person done something morally wrong? Intuitively, yes. And perhaps Morris' solution can account for this intuition (and what I would call a fact). Has this person morally wronged her (the SLRC)? Intuitively, yes. But Morris' solution doesn't account for this intuition. In order to be wronged, one must have moral standing; but this woman doesn't have moral standing according to Morris' account, and so she can't be wronged.

The first problem can be summarized as follows: Even if all SSPDs, SLRCs, children, and non-human animals have the relevant relational property and hence have moral standing, this fact (or set of facts) is contingent. Whether or not, then, a SSPD, SLRC, child, or non-human animal is wronged is contingent on (among other things) being relevantly related to others (more specifically, objects of preference). But this offends at least some persons' intuitions concerning wronging others. The intuition is that facts concerning whether someone being wronged cannot be contingent on whether that person is an object of preference of another.

The second problem is this: Look, it just is not the case that every SSPD, SLRC, child, and non-human animal is relevantly related to at least one person with primary moral standing. Our world is a cruel one, where there are plenty of beings who are not objects of preference (in the relevant way). According to Morris' account, it appears that such beings (people) do not have moral standing. Thus, in this world, there will be some SSPDs, SLRCs, children, and non-human animals that not only can't be wronged, but that our treatment of them can't be wrong. (Just think of all of the chickens, cattle, and swine that are not objects of preference. We can breed them into existence, store them away in a facility, and do what we please without doing anything wrong.) Now, if this is correct, then this should count as a weighty problem for Morris' solution. It just is the case that it's possible to do something wrong in our treatment of any, not just some, SSPDs, SLRCs, children, and non-human animals.

Why could we not do anything wrong in our treatment of these outliers who have no moral standing? Besides, we can think of cases in which we can do something wrong in our

treatment of things that have no moral standing. For example, it would be wrong to deface a sculpture, and yet the sculpture not having any moral standing. Yet this would be wrong in virtue of destroying someone's property – this would include some kind of combined ownership between the sculptor and the community of people who have commissioned the sculptor. But those outliers who have no moral standing also lack the relevant relation to other persons which could give a similar explanation for how our treatment of them could be morally wrong; in the case of outliers, if they have no moral standing, then we also can't do anything wrong in our treatment of them. For example, if a non-human animal is a pet of someone, then yes, it would be wrong to kick that pet for no reason except for the fun of it; but then again, this pet has moral standing according to Morris. What about non-human animals who are no one's pet?

Let me now discuss the merit of non-relational accounts of moral standing over relational ones. To start with, notice that just as with the first problem for Morris' account (and I would suggest all relational accounts), so too with the second problem, contingency "creeps up." Whether or not someone is an object of preference of another person is contingent. And it is not only whether or not someone is an object of preference that is contingent. So, the second problem could be summarized as: According to Morris' account, whether or not our treatment of a given outlier could be possibly wrong is contingent on whether they are an object of preference (where this then could be generalized to the contingency of whether they are relevantly related to other persons). Let's now consider a non-relational account of moral standing. Here, we would be considering some non-relational property that a being has in order for it to have moral standing. Candidate properties are: having a soul, being a human being, being alive, being rational, and being sentient. Since I think sentience is the most plausible candidate, let's use it for purposes of illustration. As we go from one possible world to another, where we change facts about objects of preference, capabilities of benefitting, and being taken advantage of, and yet we don't change the set of sentient creatures (with also the supposition that there are moral agents), I don't think that we run into the same problem as a relational account. According to a relational

account, the one and same creature will have moral standing in some possible worlds, while it will not have moral standing in others. According to the non-relational sentience account, the one and same creature will have moral standing in all possible worlds (that we consider) just so long as they are sentient.

One might object that I've cheated by holding constant the same set of sentient creatures throughout possible worlds. But this objection is spurious. I've held this constant for reasons of simplicity. We could change the set of sentient creatures. Suppose in possible worlds 1, 2, and 3, Jon is an SLRC. Suppose in possible worlds 4...n, Jon does not exist at all. Suppose further that in possible worlds 1 and 2, Jon is an object of preference (by parents and friends), while in possible world 3, Jon is not an object of preference. According to the relational account, Jon has moral standing in worlds 1 and 2, but otherwise not. According to the non-relational account, Jon has moral standing in worlds 1, 2, and 3. More specifically, according to the relational account, we can't do anything wrong to Jon in worlds 3 through n, whereas according to the non-relational sentience account, we can't do anything wrong to Jon in worlds 4 through n. And this is because Jon doesn't exist as a sentient being in worlds 4 through n.

#### *2.2.6 Conclusion with respect to Gauthier's theory*

I have introduced two problems for Gauthier's theory. These problems were the gangster-inclusion problem and the outlier problem. With respect to each, I showed how Gauthier's theory is not able to give an adequate solution to each. With respect to the gangster-inclusion problem, his theory is not able to show that the very powerful (those who I have called "the gangsters") are not able to make enforceable threats against others to get their way; as such, with such gangsters, morality as a product of the contracting process is something that runs against our shared moral intuitions. With respect to the outlier problem: Gauthier's account (with the help of Morris) is not able to show how those like SSPDs and SLRCs (as well as non-human animals) are not fully considered by the contracting process leading to the content of morality.

### 2.3 SCANLON'S KANTIAN THEORY

### 2.3.1 Characterization of contractualism

In this section, I characterize and evaluate Scanlon's Kantian theory of contractarianism, which he calls "contractualism." For now on, I will mostly refer to Scanlon's theory as contractualism. Let me preview the section as follows: In terms of the gangster-inclusion problem, I think that Scanlon's theory provides an adequate solution. The same can almost be said with respect to the outlier problem. I say "almost" because there is one such feature of Scanlon's contractualist solution to the outlier problem which I do find problematic. Scanlon's contractualism is nevertheless problematic in being inadequately suited to provide a solution to what I call the numbers problem.

What is contractualism?<sup>24</sup>

Contractualism: An action is morally permissible if and only if it is not forbidden by some principle which no one, from the set of relevant persons, could reasonably reject.

Here, two things should be pointed out. First, note that the scope of those who are considered for accepting and rejecting principles is restricted to "relevant persons." If it's possible for a person to be affected by others acting on a principle, then that person is "relevant"; they have a say in whether principle should be rejected or accepted. The second note is this: Since it is not clear what exactly "reasonable rejection" is, it will be most important to clarify what it means, and more specifically, what it means according to Scanlon.

One reason why Scanlon uses "reasonable" opposed to "rational" rejection is that rationality is understood in such a way that it may be rational for an agent to act in a way that doesn't take into account the interests of others, yet at the same time, it would be unreasonable for an agent to act in this way. He illustrates the distinction nicely in light of what could be called the "water rights example."

Water Rights Example (my paraphrase)

---

<sup>24</sup> Thomas Scanlon. *What We Owe to Each Other*. (1998).

We are to decide, for our county, principles dictating water rights. The problem is that one landowner already controls most of the water of the county, and that she has no need to cooperate with us; she is totally self-sufficient, or has the support of at least a few for her own sufficiency. Although she will most likely provide water to those who very much need it, she wants to set her own rules about the matter, allowing very little water to any given person, perhaps only at the moment when a person very much needs it.<sup>25</sup>

With respect to the handling of the issues raised in the water rights example, there is a principle allowing a continuous, yet minimal supply of water for every individual. According to Scanlon, this principle could not be reasonably rejected by the landowner, although it could be rationally rejected. (Scanlon also points out that it would not be unreasonable for everyone to propose this principle, it may nevertheless be irrational to do so if it enraged the landowner so much that it led to an outcome much worse for everyone). I agree with Scanlon on these points, and also agree that via this example, we get a better idea of what reasonable rejection amounts to.<sup>26</sup>

### *2.3.2 Contractualism and the gangster-inclusion problem*

If Scanlon's points are correct concerning the water rights example, then what does this say of his account in light of the gangster-inclusion problem and the outlier problem? With respect to the gangster-inclusion, it would be rational for the gangsters, with their combined physical and intellectually coercive strength, to reject the proposed principle that everyone have a continuous, minimal water supply, on two assumptions: The gangsters currently have a monopoly on the water, and they do not need to give the continuous supply in order to make all those who work for them maximally efficient. But here, Scanlon's contractualism is not based on rational rejection, but instead on reasonable rejection, where everyone party to the contract wants to justify to others her proposed principles, her acceptance of principles, and her rejection of principles.

---

<sup>25</sup> Scanlon, p. 192-195.

<sup>26</sup> It appears that I'm siding with Scanlon on the notion of reasonable rejection. That I am is not the case. I am, for dialectical purposes, agreeing with him about reasonable rejection as illuminated by the Water Rights Example only to eventually show problems with his account. There are others I acknowledge who disagree with reasonable rejection, and who have shown problems with such a notion.

Scanlon writes that his view “holds that thinking about right and wrong is, at the most basic level, thinking about what could be justified to others on grounds that they, if appropriately motivated, could not reasonably reject.”<sup>27</sup> The idea here is this: Gangsters, at least how I’ve described them, do not care to justify themselves to others. Instead, what is important is only their own aims and their ability to enforce threats in light of achieving those aims; they care nothing, or perhaps only inadequately care, to justify themselves – their proposed principles, their acceptance of principles, and their rejection of principles – to others. As such, gangsters would not be party, since anyone party to the contract cares to justify themselves to others. Nevertheless, even if gangsters were party to the contract, gangster-inspired principles would nevertheless be reasonably rejected. Scanlon’s account adequately solves the gangster-inclusion problem.

### *2.3.3 Contractualism and the outlier problem*

There are various outliers to the contracting process, at least on presumption. One such group includes SSPDs and SLRCs. Another group includes non-human animals. And finally, we have the group of those human beings who are non-human infants. In terms of Scanlon’s solution to the outlier problem, the four different outliers can be grouped into two categories: those who are not limited in their cognitive capacities (SSPDs) and those who are limited in their cognitive capacities (SLRCs, non-human animals, and human infants).

#### Scanlon’s contractualism accounting for SSPDs

Recall that SSPD’s were excluded from the contract not because they were incapable of understanding the terms of the contract (some of them, in fact, might be brilliant, but only in a way that is not useful), but instead because they have less to offer than what they need in return. Now, Scanlon from the start of his discussion on this problem dismisses the idea that contractualism must, or even does, think of those party to the contract as being able to benefit others in mutually advantageous way. For Scanlon, those entering into the contracting process are aiming to justify themselves to each other, independent of whether or not the being they are

---

<sup>27</sup> Scanlon, p. 15.

justifying themselves to can reciprocate in an exchange for the purpose of mutual benefit. As such, Scanlon's contractualism accounts for SSPDs in a very satisfactory way – they are directly and equally part of the contracting process, independent of the ability to reciprocate.<sup>28</sup>

Scanlon's contractualism accounting for those with limited cognitive capacities (SLRCs, and human infants and non-human animals)

What then of SLRCs? In contrast to SSPDs, SLRCs are able to offer a lot in terms of benefitting others, perhaps to the point that others could exploit their benefits for very little in return on the other hand. SLRCs are not party to the contract because they are incapable of understanding the terms of the contract; they lack the capacities for understanding the very intricate details of the contract, as well as why these details were agreed to, and most importantly for their own sake, unable to understand whether they would be ripped off by the contract.

Recall that I'm working on the assumption (a well-motivated assumption, though) that SSPDs and SLRCs have moral standing equivalent to any other person, and they have more moral standing than non-human animals. Although non-human animals are also outliers to the contract, the way to "include them in the contract" may differ drastically than how SSPDs and SLRCs are included. More importantly, if SLRCs and non-human animals are both not able to be literally party to the contract like SSPDs and other persons, then this does not mean that SLRCs and non-human animals have the same moral standing. Ignoring this point could lead to a contracting process by which SLRCs are not correctly "included" in the considerations leading to Scanlonian principles.

According to Scanlon's contractualism, all of those party to the contract are motivated to justify themselves to each other. All other things equal, SSPDs are then party to the contract. But SLRCs do not have the requisite motivation, or if they do, then they are not able to be recipients of justification, as they do not have the requisite cognitive abilities to assess justification. And non-human animals definitely do not have the motivation to justify themselves,

---

<sup>28</sup> Scanlon, p. 180.

nor can they be the recipients of justification. The contractualist solution to this is that trustees are to be introduced into the contracting process, where:

in deciding which principles could not be reasonably rejected we must take into account objections that could be raised by trustees representing creatures in this group who themselves lack the capacity to assess reasons.<sup>29</sup>

Now, in assessing the adequacy of the contractualist trustee solution, I will not point to some of the particular claims Scanlon makes concerning what he believes the trustees would object to (i.e., would reasonably reject).<sup>30</sup> Instead, I will offer two interpretations of a certain method employed by Scanlon in including outliers, specifically SLRCs and non-human animals. One interpretation, I believe, is problematic for Scanlon, the other not so much.<sup>31</sup>

---

<sup>29</sup> Scanlon, p. 183.

<sup>30</sup> On p. 183, Scanlon writes: “But not everything that interferes with animals’ living the kind of life typical of their species seems to trigger even prima facie moral objections of the kind contractualism is intended to capture. Painlessly administering birth control medication to wild animals in order to prevent their population from becoming an inconvenience to their human neighbors, for example, does not seem even prima facie objectionable in this sense (assuming that it does not cause distress to individual animals).” What Scanlon says here is questionable, to say the least. Mere inconvenience to humans does not appear to be a weighty enough reason for interference. Perhaps interference for the purpose of safer road travel, or for the flourishing of individual non-human animals (perhaps of other species even) are sufficient reasons.

<sup>31</sup> Note that the initial formulation of Scanlon’s contractualism leaves open the possibility that SLRCs be party to the contract, without the need for trustees. This is because what is fundamental to contractualism is reasonable rejection to others party to the contract of principles. We could go counterfactual here: Were it the case that SLRCs could understand the terms of possible principles of the contract, they could reasonably reject them. I find from the text very little reason why he dismisses this possibility. In fact, Scanlon writes “the idea of justifiability to them must be understood counterfactually, in terms of what they could reasonably reject if they were able to understand such a question” (Scanlon, p. 185). Here, “them” from the previous quotation refers to “young children” and “adults who do not develop normal capacities.” Scanlon continues by saying that “This makes the idea of trusteeship appropriate in their case, whether it is appropriate for the case of nonhuman animals or not” (Scanlon, pp. 185-186). From these passages, I might be able to explain why Scanlon uses the concept of trusteeship instead of counterfactual inclusion into the contracting process. Trusteeship is the simplest way to include all outliers into the contract. A counterfactual inclusion way of inclusion would be more complex, as it would allow the inclusion of young children and non-normal adults, but not non-human animals.

I think that if this is the reason to not go with the counterfactual inclusion process, then this is not a good reason at all. The process of reasoning counterfactually of course has its problems. But let’s just consider this example. Suppose that you are a rock. Suppose that you must now make decisions about what is to happen to you as a rock. I suspect this: If you are misled about what it means to be a rock, then of course, we can get counterintuitive results: For example, you might not agree to you being broken into pieces, such that you don’t exist anymore. Or you might not agree to you being grinded into sand. But this would be because you are thinking of yourself as a sentient being with interests, or that you have a soul, or that you are a self-aware being. Rocks are not sentient, do not have a soul, and are not self-aware. It doesn’t matter to them what happens to them. Thus, fully-informed and rational reasoning concerning what

### An amicable interpretation

Outliers are accounted for in the contracting process – i.e., appointed trustees – not because of any relational property they have with rational human beings, but instead because of some non-relational property they have. For example, non-human animals are appointed trustees because they are sentient and have interests. SLRCs are appointed trustees not only because they are sentient and have interests, but also because they are self-aware creatures with desires.

Scanlon says some things to suggest this interpretation. For example, Scanlon writes that “Since human beings are sentient creatures too, their pain is also something we have prima facie reason to prevent.”<sup>32</sup> So, the fact that a being is sentient gives it moral standing, and if it cannot assess reasons, then it is to be appointed a trustee. And a proposed principle in the contracting process in which upon acting on it would cause that being pain would be prima facie reasonably rejected by its trustee (although not necessarily all-things-considered reasonably rejected).

Furthermore, Scanlon writes:

Not every human being develops normal human capacities, however, so there is the question of what this criterion implies about the moral status of those severely disabled humans who never develop even the limited capacities for judgment-sensitive attitudes. The question is whether we have reason to accept the requirement to accept the requirement that our treatment of these individuals should be governed by principles that they could not reasonably reject, even though they themselves do not and will not have the capacity to understand or weigh justifications. The answer is that we clearly do.<sup>33</sup>

This could be interpreted as: As for SLRCs, it is more than their sentience which gives the moral standing. Instead, it is the fact of them having a conception of their good just like any other

---

happens to you were you a rock would amount to this: You would agree to whatever happens to you. Now, let’s extend this to non-human animals: Although as a non-human animal you are incapable of the sophisticated reasoning that many human beings are capable of, it doesn’t matter that you as a non-human animal cannot do so. What matters instead are the interests that you have. As a non-human animal, suppose a mammalian, you have an interest in prolonging your existence, in creating offspring and maintaining their existence, in experiencing pleasure and in avoiding pain. Thus, there doesn’t seem to be anything incoherent about the counterfactual inclusion idea. You are included in the contracting process by supposing that you were able to understand all that could happen to according to possible principles of the contract. Knowing your interests, you would then be able to reject or accept them. If the idea of trusteeship is not considered whacky, then I don’t see how the notion of counterfactual inclusion would be whacky.

<sup>32</sup> Scanlon, p. 181.

<sup>33</sup> Ibid., p. 185.

person, and thus are persons. So although they cannot be a part of the contracting process, they should be appointed trustees, and these trustees are to make objections to principles in light of SLRC's being full-fledged persons. In sum, beings have moral standing in virtue of some non-relational property, and SLRC's possess a non-relational property which makes them persons, and not just have moral standing like non-human animals who possess the non-relational property of sentience. Now, in summary, according to this amicable interpretation, the outlier problem is avoided. Here, we have outliers accounted for according to the interpretation just given.

#### Another interpretation

Unfortunately, there is evidence suggesting that the amicable interpretation does not fit Scanlon. Instead, I think that Scanlon ideas are not compatible to the amicable interpretation.<sup>34</sup> This is because I think Scanlon really endorses the following: Non-human animals and SLRC's do possess the non-relational property of sentience giving them moral standing. Human beings with rational capacities possess the non-relational property of being "capable of making the particular kind of judgments involved in moral reasoning."<sup>35</sup> SLRCs, on the other hand, do not have this non-relational property. So, then this invites the question? What then makes SLRCs persons? Scanlon's answer:

The mere fact that a being is "of human born" provides a strong reason for according it the same status as other humans. This has sometimes been characterized as a prejudice, called "speciesism." But is not prejudice to hold that our relation to these beings gives us reason to accept the requirement that our actions should be justifiable to them. ... The beings in question here are ones who are born to us or to others to whom we are bound by

---

<sup>34</sup> In the following sub-sub-section, I introduce a different interpretation of Scanlon, one that departs from what I've called "an amicable interpretation." This different interpretation suggests problems with Scanlon's solution of the outlier problem. But even if Scanlon doesn't endorse ideas as characterized by the amicable interpretation, I still have the problem of the fact that there is Scanlonian-like view that does endorse the amicable interpretation; this would be the view presented by the amicable interpretation. The amicable interpretation would be a Scanlonian-hybrid view within the scope of contractarian views. Because I want to show that my account is superior in being able to solve these problems (in this case, the outlier problem), the Scanlonian-like view endorsing the amicable interpretation would be equal to my account in terms of the outlier problem. All that I can say to this: Let the truth be told! If there is another account out there that can adequately solve another problem that my account can also solve, then the better! This does not mean, though, that this alternative account is a better account than mine; I need only show that it suffers from some other problem. I think that this Scanlonian-like account will nevertheless suffer from the numbers problem, as will be explained by the end of the chapter.

<sup>35</sup> Scanlon, p. 179.

the requirements of justifiability. This tie of birth gives us good reason to want to treat them “as human” despite their limited capacities.<sup>36</sup>

This passage itself involves two interpretations. I will take these interpretations in turn.

Passage interpretation 1: Perhaps the property that SLRC’s possess is, in fact, non-relational. It is “being human.” But this is an unsatisfactory necessary condition for being a person. What about intelligent alien species? What if some evolutionary process “created” a kind of being intelligent relevantly similar to human beings? Perhaps there have been intelligent beings of the past like ourselves, and then we travelled back in time and had to interact with them? Or if backward time travel makes no sense, then what if they were to travel to the future and interact with us? The questions continue. The non-relational property route invoking “being human” as a necessary condition makes little sense, fit with many counterexamples. And if it is merely meant as a sufficient condition, then we can ask why the mere fact of “humaneness” is special in some way, special in a way that elevates a human being over another sentient (yet non-rational) being.

Passage interpretation 2: The property possessed by SLRC’s is relational. They have the property of “being born to us or to others whom we are bound by the requirements of justifiability.” But as I’ve explained in relation to Morris’ Gauthierian account of moral standing, the relational property route is problematic, and for two major reasons.

Some concluding remarks concerning contractualism and the outlier problem:

Contractualism adequately accounts for SSPD’s, and I think also accounts for non-human animals very well. If the amicable interpretation is too generous to Scanlon, and something like the “another interpretation” is more fitting, then SLRC’s are not properly accounted for by contractualism. Note, though, that I don’t think that Scanlon’s trustee account is problematic.

---

<sup>36</sup> Ibid., p. 185.

Before moving on, let me just say this: I think that the alternative interpretation to the “amicable interpretation” of Scanlon is more plausible. As such, I think that Scanlon doesn’t give the fully adequate solution to the outlier problem that one would desire.

#### 2.3.4 *The numbers problem*

In this subsection, I characterize a problem which draws on a certain issue already familiar to many working in normative ethics. This problem is not one which is uniquely proposed for contractarian accounts, but instead is one which any developed ethical theory (theory of permissible/right action) should be evaluated with respect to.

Suppose that you have a choice between saving one person from great harm of intensity,  $x$ , and five persons from each experiencing great harm of intensity,  $x$ . To make things more concrete, suppose that you have a boat, and you can either rescue a group of one person from drowning or you can rescue a group of five persons from drowning, but you cannot rescue both groups. If you save the one, the five will drown. If you save the five, the one will drown.<sup>37</sup> Many people will have the intuition that it is a better state of affairs that the five are saved, thereby only one dying, than the state of affairs in which the one is saved, and the other five die. Many people have the intuition that you are morally required to save the five over the one, and even more people have the intuition that you are at least morally permitted to save the five over the one.

Now, I will state the numbers problem:

The numbers problem: An adequate ethical theory must be able to show that, other things being equal (and absent considerations of rights and desert), it is at least

---

<sup>37</sup> The numbers problem relies on the fact that many of us share the intuition that when all other things are equal, and there is a choice that needs to be made between saving one and saving five, it is morally permissible to save the five. Yet this intuition can of course be questioned. To see how one might question such an intuition, see John Taurek’s “Should the Numbers Count?” *Philosophy and Public Affairs*. 1977. And as for replies and other relevant material, see Samuel Scheffler’s *Rejection of Consequentialism*. Clarendon Press; Oxford University Press: 1982. Also, see Sophia Reibetanz’s “Contractualism and Aggregation” *Ethics*. 1998. Finally, consult Derek Parfit. “Innumerate Ethics.” *Philosophy and Public Affairs*. 1978.

permissible for an agent to save five people from each suffering harm, of intensity  $x$ , than save the one person from suffering harm of intensity  $x$ .

### *2.3.5 Contractualism and the numbers problem*

In this subsection, I argue that contractualism cannot adequately account for the numbers problem. Given how Scanlon proposes contractualism, his account does not solve the numbers problem. I make my case, in outline, as follows. First, I actually suppose that Scanlon's account gets us the intuitive result in numbers problem-style cases. But the rationale that allows his account to do so then also leads to counterintuitive results concerning other cases. As such, we should reject this rationale. Thus, we should also reject this way of going about handling the numbers problem. In other words, this route of handling the numbers problem is not a solution to it (as following it will lead to counterintuitive results in other cases). Second, I argue directly that Scanlon's account does not solve the numbers problem. My showing this depends on fleshing out the concept of reasonable rejection, a concept that is central/fundamental to Scanlon's contractualism. Before getting into the two modes of rejecting his contractualism in light of the numbers problem, it is important that I first discuss Scanlon's attempt to solve it (of course).

Scanlon uses a tie-breaker argument to show that it is permissible to save the five over the one. There are two main steps to the argument. First step: Show that there is motivation for a principle in which an agent is required to save the five over the one. Second step: Show that this principle does not suffer from weighty enough objections for it to be reasonably rejected.

#### First step

The principle requiring one to save the greater number is motivated by assuming that there is a principle permitting one to save the fewer number – i.e., the one and not the five.

Scanlon writes:

In such a case, either member of the larger group might complain that this principle does not take account of the value of saving his life, since it permits the agent to decide what

to do in the very same way that it would have permitted had he not been present at all, and there was one person in each group.<sup>38</sup>

By acting on the principle (permitting one to save only the one), one would then be acting in a way that did not recognize the other people of the group of five. It's as if the value of each of the four of the group of five "makes no difference to what the agent is required to do or to how she is required to go about deciding what to do."<sup>39</sup> This principle of permissibility motivates then the proposal of a different principle, namely one requiring that an agent save the five over the one – more generally, saving the greater number.<sup>40</sup>

### Second step

Now we consider what objections could be made to this principle of requirement. One objection: The person in the group of one says that she is not being accounted for in the same way as the others in the group of five are accounted for. This objection has no force since, indeed, her presence, and her value as a person, was indeed accounted for with the same weight as each of the five. Second objection: The person in the group of one says that she would have been better off under the principle permitting the agent to save the group of one, rather than under the principle of requirement. In fact, in light of this, she is made worse off by an agent acting under the principle requiring the saving of the five over one. But this objection is no good according to Scanlon, as he writes:

It will be true of almost any principle that someone would have been better off if some other principle were in effect. And, as we have seen, these particular alternatives are subject to a strong objection that does not apply to the one now under consideration.<sup>41</sup>

---

<sup>38</sup> Scanlon, p. 232.

<sup>39</sup> Ibid., p. 232.

<sup>40</sup> Note that I use the language of "proposal of principles." But with Scanlon's contractualism there really need not be any kind of "proposing" of principles. Instead, we could imagine a set of candidate principles of morality. From this set, we (or: the contracting agents) ask the question: Of these candidate principles, which are the ones that cannot be reasonably rejected? Those candidate principles which cannot be reasonably rejected are the principles of morality. Now, I use the language of "proposing" because it is easier to explain the process of reasonable rejection by use of this term (and the concepts associated with it).

<sup>41</sup> Ibid., p. 233.

So, the person in the group of one cannot reasonably reject the principle on the basis that an alternative would have made her better off, and furthermore, that the principle in fact makes her worse off.

### The complaint model and problems with the second step

I will now explain why I think that Scanlon's argument is problematic, in particular the second step. Scanlon endorses a certain characterization of his contractualism, with a few exceptions. He writes:

All the grounds for rejecting a principle that I have so far considered arise from generic reasons that an individual would have who occupied a certain position in the situations to which that principle applies. This suggests what Parfit has called the Complaint Model. On this interpretation of contractualism, a person's complaint against a principle must have to do with its effects on him or her, and someone can reasonably reject a principle if there is some alternative to which no other person has a complaint that is as strong.<sup>42</sup>

Now, Scanlon does note two departures he takes from the complaint model characterized above. Both of these departures, though, are irrelevant to the criticism I will offer. I want to point out that the complaint model involves two claims: (i) a complaint is made by an individual and must refer to the effects on herself; (ii) a sufficient condition for reasonable rejection of a principle is that there is some alternative by which no one else has a complaint/objection at least as strong. Thus, (ii) is not a necessary condition for reasonable rejection, and furthermore this suggests that there are other sufficient conditions other than (ii) for reasonable rejection. Given the plausibility of (i), I think that we can construct from our commonsense usage of "reasonable rejection" at least another sufficient condition – apart from (ii) - for reasonable rejection of principles.

To start, reasonable rejection is made by individuals. But of course, the reasonability of a complaint/rejection does have to do with what would happen to others under alternative principles, as evidenced by Scanlon's endorsement of (ii). The question is: How far should we go in considering others in making a reasonable complaint/rejection? Of course, if a principle requires that another person be prevented from suffering losing both her legs with the

---

<sup>42</sup> Scanlon, p. 229.

consequence that I am not saved from a minor cut on my face, then I, despite how much I care about myself, cannot reasonably reject the principle. Also, let's consider ties between groups in *minor* harms. Suppose that there is a principle requiring that a group of persons each be spared from minor cuts on their faces, with the consequence that that I am not spared a cut on my face. Then I could not reasonably reject this principle. With respect to ties, I think the same thing can be said about minor benefits. On the supposition that I am minimally well-fed and the other five are minimally well-fed, consider a principle that requires that five Italian-cuisine meals be delivered each to five persons, with the consequence that I do not have the same cuisine meal delivered to me. I could not reasonably reject this principle.

Suppose that the tie-breaker argument is successful in getting the correct result (i.e., the one that is reflected by our shared intuitions) with respect to saving five opposed to saving one (and more specifically, that it is at least permissible to save the five over the one). Then we should think about how this tie-breaker argument would be used to assess other cases. It would (most likely) get the correct result in the Trolley case (that it's permissible to kill one in order to save five). But we should ask then: Would it get the correct result with respect to the Transplant case? (Recall that the correct result, at least the one reflected by our shared intuitions, is that it is impermissible for the surgeon to proceed to kill and to transplant organs to save the five.) I will argue that it doesn't. To give a preliminary summary: If we can suppose that Scanlon's contractualism gets the correct result in the numbers problem via the tie-breaker argument, then we will see that the same reasoning leading to this also leads to the permissibility of the surgeon proceeding in the Transplant case. Yet this is a counterintuitive result.

Let's expand. On the reasoning that would make it permissible to save the five over the one according to the tie-breaker argument, we could imagine the second (through the fifth) person of the group of the five needing a transplant that the interests of the one healthy person (and possible donor) has already been taken into consideration. As such, the second (through the fifth) person could complain that their interests were not taken into consideration if the matter was

settled at that. The second (through the fifth) could say that the possible donor's interests had already been taken into place by the one-to-one comparison between her and the first person in the group of ailing (due to organ failure) persons.

It is at this point that Scanlon might mention rights in such a way to explain why in the trolley case and numbers problem cases it is permissible to save the five over the one, while in the transplant case, it is impermissible. He might say that rights are not violated in the trolley case and the numbers problem cases by saving the greater over the fewer, while in the transplant case, the saving of the greater over the fewer involves a violation of rights. Indeed, Scanlon does allow for matters of rights and entitlement to play a role in what is reasonably rejectable. But I think that this maneuver would be circular: An agent has a right against being treated in a phi-way just in case this agent has a right against being treated in phi-way.

To avoid this charge of circularity, I suspect the following: Look, it's just the case that we have certain rights. Justifying principles of action to ourselves and others requires that we reflect to ourselves and others what these rights amount to. In the case of Transplant, we have a case in which we each have a right to not be used in this way, if we were one of the possible donors. Thus, it is impermissible for the surgeon to proceed to kill us and transplant our organs for the purpose of saving five others. My reply to this way of understanding this way of understanding rights in relation to reasonable rejection is this: If this is the way to reason, then it will turn out that pulling the lever in Trolley is impermissible. According to the above reason, what prevents the one person tied to track B (the alternative track) from being permissibly killed? According to the argument presented above, this one person in Trolley has a right not to be killed in this case. Therefore, with all other things equal, it is impermissible to kill her. And yet the shared intuition is that it is permissible to kill her (for the purpose of saving the five).

Now, at this point, I will change directions. I will argue that the tie-breaker argument doesn't lead to the correct result, in the cases relevant to the numbers problem. My argument is this: Recall that reasonable rejection is relativized at least partly to the individual who gives the

rejection. Although it is not the case that the mere fact of rejection from an agent counts as reasonable (e.g., think of cases in which an individual rejects a principle that requires her to pay \$10 in order to help 10 people from dying), there are other cases in which rejection of a principle on part of an agent we would think is reasonable, even if it is the case that this principle is correct. This last “factoid” is something that plays a crucial role in showing that Scanlon’s contractualism, relying on reasonable rejection, cannot solve the numbers problem.

In the case of the transplant case, we can easily see that the one person who might be the possible donor is also one who could reasonably reject a principle allowing for her (or anyone healthy) to be killed and used for transplant to save five others. As such, according to Scanlon’s contractualism, we get the intuitive result that it is impermissible to save the five dying from organ failure by killing the one healthy person for the purpose of organ transplantation into the one. The impermissibility of doing as such is determined by the fact of reasonable rejectability by the one person who is possibly the donor in this case. Yet if this were the case, then we could imagine reasonable rejection on the part of the one person tied to the alternative track in the trolley case. Put rhetorically: Wouldn’t this one person reject a principle permitting the switching of the lever that would save five, yet would kill him/her? Well, the answer, I think, is yes. Now, let’s consider the numbers problem. Here, we could imagine the one in the group of one rejecting a principle saving the five over the one. More specifically, this principle, if acted upon, would mean that the one would be left to die and yet the other would be saved from dying. If one were the person that would die if the other five are saved, then I think that we can safely say this: A principle requiring this would be reasonably rejected by the one. In other words, if a principle deeming killing the healthy person for organ-transplant to save five dying from organ failure is reasonably rejectable, then so too should principles deeming pulling the lever in the trolley case or in saving five versus saving one (numbers problem cases). Yet we think that it is permissible to pull the lever (and thereby kill one in order to save one), as well as think that

saving five over one is permissible. According to the reasonable rejection account, this would be impermissible. Therefore, the reasonable rejection account, central to contractualism, is false.

## 2.4 CONCLUSION

In section 2.2, I examined Gauthier's Hobbesian style theory. More specifically, I evaluated its plausibility in terms of two problems. The first problem was the gangster-inclusion problem and the second was the outlier problem. With both problems, I argued that his account does not offer an adequate solution to these problems. In section 2.3, I examined Scanlon's Kantian style theory. Whereas he gives a very adequate solution to the gangster-inclusion problem, and he gives a solution to the outlier problem which, under certain interpretations, is very adequate, his account nevertheless suffers from what I call the numbers problem. Central to his account is the notion of reasonable rejection. According to how I've analyzed this notion, Scanlon does not provide an adequate solution to the numbers problem.

## CHAPTER 3 – ACT VERSUS RULE CONTRACTARIANISM

### 3.1 INTRODUCTION

In this chapter, I argue for an act-based view of contractarianism as an ethical theory, as opposed to a rule-based view. More specifically, I argue that the contracting agents do not make decisions about what rules or principles to agree to, but instead that they decide directly on what actions to agree to. As far as I can tell, no such theory has been developed.<sup>43</sup> It is in this chapter that I defend such a view.

Recall that my account is one endorsing the veil of ignorance of Rawls' contractarian theory. As formulated in chapter 1, by combining the Rawlsian component and the act component, I arrive at the following ethical theory:

Rawlsian ethical act contractarianism: An action A is morally permissible if and only if, from behind a veil of ignorance, A would be agreed to by all contracting agents, each of which are the way they actually are in real life, except that they are perfect expected individual utility maximizers.

In the next chapter I defend my characterization of the contracting agents as being placed behind a veil of ignorance and employing the principle of maximizing expected individual utility. In this

---

<sup>43</sup> Such a view is mentioned explicitly, but not defended, by Shelly Kagan in his *Normative Ethics*. Westview Press: 1998. Consult pp. 240-243. Kagan distinguishes between rule and act contractarianism, where what he focuses on in making the distinction is the “focal point” of what is evaluated by the contracting agents; for rule contractarianism it is rules, whereas for act contractarianism it is acts. It is important to point out, though, that Kagan does not give a defense of one kind of contractarianism over the other. In fact, he illuminates contractarianism in terms of its rule-form, since it is the one that is better known to audiences. Also, Judith Jarvis Thomson discusses something like an act contractarian position in her *Realm of Rights*. Harvard University Press: 1990. See pp. 176-197. In particular, Thomson is discussing the Trolley case, in which what is to be decided is whether it is permissible for a person to pull the lever under certain specified circumstances. She then asks whether “the workmen” involved in the situation would consent to the lever being pulled. Here, what is being evaluated is not a rule, but instead a specific action, namely that of pulling a lever. Nevertheless, Thomson rejects hypothetical consent (and the project of contractarianism in general) when she concludes: “So hypothetical consent is not only not sufficient for permissibility, it is not necessary for it either” (Thomson, p. 188).

chapter, I'm interested in defending the act component. Nevertheless, my reader will see that discussion of the veil will be relevant in motivating the act component of my account.

Now, one might ask what I mean by an action "being agreed to by all contracting agents." As noted in chapter 1, the notion of agreeing to an action seems mysterious. But really it is not anymore mysterious than agreeing to a rule. Consider again rule contractarianism. In the case of morality, the contracting agents would consider from a list of rules which ones to agree to in regulating their behavior, or whether to not agree to any such rules. So, there is the notion of a non-agreement point which must be specified according to contractarian accounts. That is, what will happen were the contracting agents not to come to an agreement about the rules? As I discussed in chapter 2, without a veil, contracting agents with lesser capacities and powers will be placed at a disadvantage and may agree to unpleasant terms, terms which are nevertheless preferable to the non-agreement point of "might makes right", where the stronger have their way in any situation. With a veil, of course, a contracting agent does not know her own capacities, powers, and preferences. As such, it is more plausible to think that the contracting agents will reach an agreement to more pleasant terms/rules for everyone.

With respect to act contractarianism, the notion of a non-agreement point is also relevant. I should note, though, that unlike rule contractarianism, the contracting agents of act contractarianism are examining actions in a case-by-case manner. Relativized to a specific case, the contracting agents consider what action or actions are to be agreed to with respect to the acting agent in the case. But the idea of a non-agreement point also enters into the reasoning of the contracting agents. If they agree to an action, then they have compared the outcome of what would happen were the acting agent to perform that action with what the acting agent could otherwise do were no action agreed to. Let me illustrate this with an example. Take a very simple case of aid, for example the shallow pond case, where an adult passing by a shallow pond witnesses a child drowning in a shallow pond and the only way to save the child is by the adult getting her pants wet in order to save the child. Non-agreement in this case would amount to it

being permissible (but not required) that the adult not save the child. I suspect that a large majority of contracting agents (absent a veil) would agree to the action of the adult saving the child, and *only* that action (relativized to the case). The contracting agents would compare (i) the outcome of the adult acting according to the action of saving with (ii) the adult permissibly doing otherwise upon there being no agreement. And most contracting agents (absent a veil) would agree to the action of saving upon making this comparison. Notice, though, that agreeing to an action does not actually determine how agents will, in fact, act. Just because the contracting agents agree to an action does not mean that real-world agents will act that way. And just because the contracting agent do not come to an agreement does not mean that real-world agents will act according to what is now permitted according to the non-agreement point.

Agreement to an action by the contracting agents amounts to them agreeing to the action that if performed is preferred (based on outcomes) over other actions if they were performed or actions permissibly performed upon there being no agreement whatsoever. It is open, though, from one act contractarian to another just exactly what is agreed upon based on the rational choice principle endorsed. In the case of my account, the contracting agents are placed behind a veil of ignorance. Since they are placed in a situation of uncertainty via being placed behind a veil of ignorance, they make their decisions in light of the principle of maximizing expected individual utility. At this point, it is relevant to discuss the role the veil of ignorance plays in determining which rational choice principle the contracting agents use in order to come to agreement (if they come to agreement at all). (I should note, though, that since they are placed behind a veil of ignorance, the decision made by one contracting agent will be identical to the decision of any other contracting agent. As such, agreement will be reached, and unanimous agreement by multiple agents will be identical to a decision made by one.)

In discussing the role of the veil in agreeing to actions, it makes sense to discuss in more detail what may occur without a veil. Without the veil, it is implausible to think that they would make agreements according to maximizing societal (or overall) expected utility. Because each

would know her position with respect to the available alternatives, and because there may be cases in which maximizing societal expected utility does not maximize utility for some individuals, it would be implausible to think that the contracting agents would go along and make decisions in this fashion. Now, suppose without a veil, the contracting agents would make decisions based on maximizing individual (and not overall or societal) expected utility. But if this were the case, then agreement would not always occur. Some actions will maximize individual utility for some while not do so for others. All other things equal, there would not be an incentive to agree upon an action which does not maximize individual utility. Thus, one can see that I'm making the claim that it is only by being placed behind a veil of ignorance that the contracting agents will make agreements according to maximizing expected utility, and more specifically maximizing expected individual utility.

In order to furthermore contrast my view with typical rule-based contractarian views, I should mention the difference in the "directness" of each view. We can contrast the two theories at the most basic level as follows:

Act contractarianism: An action, A, is morally permissible if and only if A is agreed upon by the contracting agents under the relevant conditions.<sup>44</sup>

Rule contractarianism: An action, A, is morally permissible if and only if A is not prohibited by any moral rule agreed upon by the contracting agents under the relevant conditions.<sup>45</sup>

---

<sup>44</sup> Note that there are various ways of spelling out what it means to agree to an action according to an act contractarianism. For my view, agreement upon an action amounts to it being the one that maximizes expected utility of its alternatives. But an act contractarian view need not endorse this way of understanding agreement to an action. One could endorse a different kind of rational choice principle for the contracting agents, and thereby endorse a different kind of property that an action has when it is agreed upon; perhaps the action agreed upon is one that satisfies the maximin principle. Also, note that if there is no veil of ignorance, this can change things quite a bit. Perhaps with no veil of ignorance, the contracting agents will agree to the action which makes no one worse off. Or, as I've suggested at other places in the dissertation, agreeing to an action (when there is no veil) amounts to agreeing to the action which the stronger contracting agents can persuade the others to.

<sup>45</sup> Note that a formulation of rule contractarianism could replace the mention of "rules" with "principles."

So one can see that act contractarianism's evaluation of actions is direct as opposed to the indirectness of rule contractarianism. The direct objects of agreement for rule contractarianism are rules, which then apply in the evaluation of actions; the direct objects of agreement for act contractarianism are actions.<sup>46</sup>

In what follows, the first thing that I will do is explain a certain kind of measure (called THE MORAL RULE THESIS) I'll employ in determining whether a rule-based theory is a genuine alternative to the kind of theory I propose, namely an act-based one. Next, I explain in more detail how an act-based contractarian view works. I then explore the act versus rule utilitarianism debate in order to see if parallel argumentation could be used in arguing for act contractarianism over rule contractarianism. I ultimately conclude that rule contractarianism either violates a constraint on what moral rules are (where such a constraint is backed by our ordinary understanding of moral rules) or it is implausible. (This last comment about implausibility actually oversimplifies the nature of the charge, but at this point, it suffices to just say that rule contractarianism runs into the problem of being implausible.)

### 3.2 THE MORAL RULE THESIS

Because I'm arguing against rule contractarianism, it is important that I give some kind of standard for what counts as a moral rule. Moral rules are supposed to be simple enough to understand, remember, and to be able to apply at will. These are practical constraints on what a moral rule can be. If a rule is too long or complicated that it cannot be understood or recalled or applied at will by normal adult human beings, then we would not count it as a genuine moral rule.

---

<sup>46</sup> Because I'm advancing a controversial element to an ethical theory, I should say something of various other elements that are thought to be accounted for by an ethical theory. My contractarian account is concerned primarily with the permissibility of actions. But what of such things as rights, duties, and virtues? More specifically, there are those who think that the permissibility of actions is more fundamentally explicable in terms of the rights of persons involved in moral cases. Others think that duties are fundamental, and there are some who believe virtues to be fundamental. For example, whether or not it is permissible for a pregnant woman to abort will be determined by what rights the pregnant woman, fetus, and any other relevant person have. According to my account, rules, principles, rights, duties, and virtues are not fundamental in this way in evaluating actions. Talk of rule, principles, rights, duties, and virtues comes after determinations of permissibility. I understand that this claim has no defense at this point, nor is it my goal to give a strong defense to such a claim. Instead, I simply note this issue, and what kind of (undefended) position I take with respect to it.

This does not exclude the possibility of moral rules having exceptions built into them. Yet with built-in exception clauses, there are certainly limitations placed on just how much exception is allowed in order to maintain simplicity and avoid violations of practicality constraints. So, imagine a rule which was something like 50,000 words long, which took into all considerations such that when applied to any given moral case would render the intuitive result. Such a rule would no longer be counted as a moral rule according to our ordinary understanding.

So, I propose the following thesis to capture the kind of practical constraints on moral rules:

THE MORAL RULE THESIS: Moral rules are to be short and simple enough so that they can be effectively applied by all normal adult human beings.

Now, this thesis leaves open some vagueness. We could ask – Is a rule with 41 words short and simple enough to be effectively applied by any normal adult human being? I don't know the answer to this. But I think most of us would say that, all other things equal, a rule of 12 words will be short enough and not violate the constraints imposed by THE MORAL RULE THESIS, whereas a rule of 52 words will most likely violate such constraints.

Now, if a rule-based theory advances moral rules which violate THE MORAL RULE THESIS, then this is indicative of such a view not really being a genuine alternative to an act-based view. Recall that the basis for violating the thesis would be that the rules would no longer be short and simple, and as such, would no longer count as practical rules. This point will be relevant in section 3.4, which deals with the possible practical equivalence between rule and act contractarianism.

### 3.3 ILLUSTRATION OF HOW THE ACT COMPONENT WORKS

In this very brief section, I illustrate how the act component of my account works. The question may arise about how exactly an ethical theory is to proceed if there are no moral rules/principles it endorses. According to my account, the outcome of contractual agreement is not rules/principles but instead decisions concerning actions in specific moral cases. Although I've already talked about how the contracting agents agree upon an action in light of it being the one

of its alternatives which maximizes expected utility, it is still not clear to the reader just exactly how this account works with respect to specific moral cases. In this section, I clarify the account in light of this quandary.

The best I can do in explaining my view is by illustrating an example. Since I will make reference to this example throughout the paper, let me call it THE CONFLICT CASE. The case is as follows: Suppose a person A makes a promise to meet with person B the next day at noon. As he's walking to his meeting, at 11:15 a.m. he receives a cell phone call from person C. C says that it is urgent, that she is in quite a bit of distress, and needs him to meet with her at 11:30 a.m. The meeting with C will take at least an hour, and B does not have a cell phone to be contacted at. A agrees to meet with C at 11:30 a.m. As A is walking to meet with C, A sees that a bicyclist has had an accident two blocks ahead. It doesn't look greatly serious, but injury does look to be the case, and the accident has happened on a remote street. Person A can foresee that going up to the bicyclist and involving himself will probably be at least a half an hour ordeal; this will make A be late for his meeting with his distressed friend. A decides to run toward the bicyclist. But then as he's running down the street, he sees a child drowning in a pond in the park adjacent to the street. If he runs to rescue the child, he will not be able to help the bicyclist. A decides to rescue the child.

I suspect that most people would judge that A has done what is morally permissible, and that doing otherwise would be morally impermissible. According to any plausible act contractarian account, the contracting agents would decide that A's rescuing the child is permissible, and that doing otherwise is impermissible. They would be presented with the entire case before making their decision. And most importantly, they are placed behind a veil of ignorance. Given their placement behind a veil, they will reason according to the principle of maximizing expected utility as the rational choice principle for making their decisions and coming to agreement about actions. Contra Rawls, probability assignments are made according to the principle of insufficient reason, and utilities are assigned according to the kinds of

preferences that the various agents have. Applying the principle of maximizing expected utility, the contracting agents would arrive at the decision that aligns with our shared intuitions – in this case, that A’s rescuing is required. I will now illustrate how they arrive at such plausible results in light of this case.

To further explain THE CONFLICT CASE, there are five people involved, with one of them being the acting agent.<sup>47</sup> Let’s suppose then that there are four people whose well-being, or preferences, should be taken into consideration. (In a sense, the four are moral patients, where the other is the moral agent, and so we will only make utility assignments for the four moral patients.) Now, because of the complexity of this case, in a sense, there are four available actions that can be assessed. (See below the matrix.) The question of whether an agent ought to in the end save the drowning child, thereby foregoing helping anyone else, can be settled in various ways. I will explain my favored way of settling the issue, and merely footnote another way.<sup>48</sup>

---

<sup>47</sup> For the purpose of clear exposition of the following example, I will only consider the well-being of the four persons who are affected by the moral agent’s actions. Now, it is quite plausible to consider the well-being of the agent acting. Certainly, there are many cases in which the agent, in acting, is affected in a way that either benefits or harms her own well-being. As such, this is relevant to consider in assigning utilities. In fact, my account, one that employs a veil of ignorance, makes sense of this fact. But for the purposes of clearly illuminating the example, I will withdraw consideration of how the agent is affected. Were it the case that I include the well-being of the agent acting, then this will not affect, in this case, results concerning the permissibility of her actions (although of course this is a contingent matter).

<sup>48</sup> According to this method, we assess the four available actions step by step. For example, the contracting agents could evaluate the first two actions available to the agent according to the first step: Should the agent keep the appointment with A or should she not do this, instead meeting with her distressed friend? Note in the matrix below, the contracting agents would be considering only rows (i) and (ii), and columns (I)-(II). Note also that I’ve made the utility assignments, but in effect what I have done is considered the preferences of each person A and B, and have made further assumptions what a “normal” person would prefer. Because it is equally likely to be in the position of A as it is in the position of B, the contracting agents may ignore probabilities, and then just add up the utilities with respect to each of the two actions. Because the expected utility of keeping appointment is less than the expected utility of meeting distressed B, the contracting agents would decide that meeting the distressed B is required. Step two: The contracting agents are to only consider rows (ii) and (iii) and columns (II) and (III). Here, the preferences of B and the bicyclist are considered in making utility assignments. And since there is greater expected utility to helping the bicyclist, this is the action which is required. So, although by the first step meeting with distressed B is required, by the second step it is now impermissible. The third step considers only rows (iii) and (iv) and columns (III) and (IV). Once again, utility assignments are made and we see that saving the drowning child maximizes expected utility. Since there is no further available action, the third step is the terminal step. According to the contracting agents, the required thing to do in the end is to save the drowning child, where doing anything else is impermissible.

Because it's the case that the agent, if she has followed the dictates of morality, will have to ultimately decide between the bicyclist and the drowning child, the most direct way of the contracting agents to proceed in making their decision is by assigning utilities to columns of I-IV, for rows (i)-(iv). They then sum up each of the rows in order to make total expected utility assignments for column V.<sup>49</sup> The action which has the greatest total expected utility in column V will be the one which is decided to be required according to the contracting agents. They will have agreed upon that action because it is the one that maximizes expected utility. Note then that the action which maximizes expected utility is saving the drowning child, thereby not helping the injured bicyclist, not meeting with distressed B, and not keeping the appointment with A.

		I	II	III	IV	V
		A	B	bicyclist	child	Total
i	keep appointment with A	5	-10	-20	-100	-125
ii	meet distressed B	-5	10	-20	-100	-115
iii	help injured bicyclist	-5	-10	20	-100	-95
iv	save drowning child	-5	-10	-20	100	65

In concluding this section, I should now note how a rule contractarian account, in contrast, would have to approach THE CONFLICT CASE. According to a rule contractarian account, in order to evaluate what one ought to do in the example above, there needs be a rule. But as I will argue later in this paper, there is a problem for rule contractarians: Such a rule to dictate what one ought to do in such circumstances amounts to a violation of THE MORAL RULE THESIS as laid out above. More specifically, in order to obtain plausible evaluative results concerning cases like the example provided above, a rule contractarian approach will require there to be rules that violate THE MORAL RULE THESIS. Once again, all of this will be argued for in detail in what follows.

---

<sup>49</sup> It is important that I point out that the contracting agents making decisions which maximize expected overall utility comes about by their attempting to maximizing expected individual utility. This is because in being placed behind a veil of ignorance, they do not know what position they occupy under any of the various alternatives of action. Under any outcome, they do not know what position they will occupy. As such, they reason according to the principle of maximizing expected individual utility; it just so happens that in making decisions according to this principle, they will also maximize expected overall utility.

### 3.4 PRACTICAL EQUIVALENCE

In this section, I examine the nature of the practical equivalence objection which can be leveled against rule contractarianism. In doing so, I will make mention of some of the debate that occurs between rule and act utilitarianism. I will nevertheless not discuss the debate in detail, and instead will discuss the debate only where needed to illuminate significant points with respect to act and rule contractarianism.

I should first point out that the charge against rule utilitarianism (or, more broadly: rule consequentialism) is usually construed as a dilemma: Either rule utilitarianism avoids the charge of incoherence at the cost of being practically equivalent to act utilitarianism, or it avoids the charge of practical equivalence at the cost of being incoherent. In this section, I will explain both (i) how rule contractarianism arrives at a practically equivalent account to act contractarianism and (ii) why practical equivalence is a charge against rule contractarianism. The nature of incoherence, on the other hand, should be understood as what has been called “rule-worship.” Now, it is important to point out that the alternative to rule contractarianism’s practical equivalence is *not* incoherence/rule-worship. Instead, the alternative is to endorse a view of moral rules which is either (i) indeterminate, (ii) implausible/counter-intuitive, or (iii) incoherent/rule-worship based. I will explain the charges of (i)-(iii) in section 3.5.

Because this section concerns practical equivalence, it only makes sense to characterize what it is: Two or more theories are practically equivalent just in case they require an agent to perform the very same action, for all possible actions; if an action is morally permissible according to one theory, then it is morally permissible according to the other(s).<sup>50</sup> Now, recall

---

<sup>50</sup> In the case of rule utilitarianism, one might think that practical equivalence is problematic for this reason: If rule utilitarianism is practically equivalent to act utilitarianism, then it would collapse into a theory that advocates actions which are very much counterintuitive. For example, it would in the end advocate actions which violate rights, or which ignore obligations we have to persons we share special relationships with, or which are way too demanding. Of course, some rule utilitarians may deny this. And they may deny this on a couple of grounds. First, they might just deny that in cases where we would intuitively say that rights are violated that anything morally wrong is being done. Or, they may say that a more sophisticated act utilitarianism to which rule utilitarianism is practically equivalent would not allow for rights violations, the

THE MORAL RULE THESIS. If a rule-based view, such as rule utilitarianism, ends up being practically equivalent to an act-based view, such as act utilitarianism, then this will come about because the rules advanced by the theory are so long and complicated that they violate THE MORAL RULE THESIS. The same goes for rule contractarianism. A rule-based view, such as rule contractarianism, maintaining practical equivalence will advance rules which are not practical moral rules, and thus the theory itself is not a genuine alternative to its act-based counterpart theory. Indeed, the charge of maintaining practical equivalence only so that one's theory is not even a genuine alternative is problematic for the theory.

### 3.4.1 *Practical equivalence*

In defending his own brand of rule utilitarianism, Hooker notes three versions of the practical equivalence objection leveled against such a theory. The most sophisticated version of the objection, at least in my eyes, is as follows. Rule utilitarians have to get clear about how “exceptionless” their rules are.<sup>51</sup> If they allow for exceptions, then they will want to explain what motivates these exceptions. For example, if a rule like “Don’t steal” is modified to “Don’t steal except in cases with such-and-such properties”, then we should ask why cases with such-and-such properties count as legitimate exceptions. If the reason is because those are cases in which stealing maximizes overall good, then indeed rule utilitarianism is practically equivalent to act utilitarianism. In following rules with such exceptions, one would be doing what is overall best – i.e., one would be doing what act utilitarianism dictates.

And we can see how specific the exceptions could be, in order to accommodate for cases in which deviating from the rule would produce more overall goodness. To take an example from

---

ignoring of special obligations, or demandingness. As such, rule utilitarianism’s practical equivalence would not be problematic at all. The reason I don’t think that this is the real reason why practical equivalence is problematic for rule utilitarianism has to do with the commitments one would have in endorsing any form of utilitarianism. The charge that a rule utilitarian account ends up endorsing actions which run counter to our shared intuitions is not much of a charge for someone who is already committed to utilitarianism. They will just bite the bullet with respect to this charge, saying that running counter to such intuitions is not indicative of the theory being false.

<sup>51</sup> Much of what I say concerning the debate between act and rule utilitarians, and how that debate unfolds, comes from Brad Hooker’s “Rule Consequentialism.” *Stanford Encyclopedia of Philosophy*, and also Hooker’s *Ideal Code, Real World*. (2000).

Hooker, suppose that the rule is “Don’t break promises *except when breaking a promise to meet one person will enable you to meet someone else who will benefit at least slightly more.*”<sup>52</sup> Any specific exception that if acted upon would cause greater goodness could be maintained by the rule utilitarian. But in doing so, rule utilitarianism would be practically equivalent to act utilitarianism.

In light of the above-mentioned rule, this kind of practical equivalence is not problematic. It certainly coheres with the practical constraints of THE MORAL RULE THESIS; the rule is simple and short enough in order to be effectively applied to any given specific moral case. But we can amplify such exception clause based rules by reconsidering THE CONFLICT CASE. In other words, we can show that practical equivalence will require that the rules are so long and complicated that they violate THE MORAL RULE THESIS. In the next subsection, I will show that in light of THE CONFLICT CASE, the contracting agents will agree to rules which violate THE MORAL RULE THESIS.

#### *3.4.2 Practical equivalence for rule contractarianism*

In this section, I explain in more detail how a rule contractarian view would maintain practical equivalence. Of course, one might wonder at this point why a rule contractarian view would want to maintain practical equivalence if in doing so it would come at costs to such a theory. The reason that a rule-based view, such as rule contractarian views, would want to be practically equivalent has to do with a corresponding motivation to arrive at results in specific moral cases which align with our shared intuitions. To illustrate this point, consider again THE CONFLICT CASE, there were numerable circumstances surrounding what person A ought to do. Recall that the decision was that it was permissible for the acting agent to save the child drowning thereby not being able to help the injured bicyclist. The result of saving the child was not only not helping the bicyclist, but also breaking his promise to meet on time with his distressed friend and also breaking his promise to meet his other friend for lunch. And, in fact, we may say that doing

---

<sup>52</sup> Brad Hooker. *Ideal Code, Real World.* p. 96.

otherwise would have been impermissible for the acting agent. I showed how an act-based view would arrive at this result. Now, a rule-based view would similarly want to arrive at this result. And if we were to continue adding moral cases in which act contractarianism got the right result, then rule contractarianism would similarly be motivated to get the right result. Here, we would have a rule-based view which is practically equivalent to its act-based counterpart theory. So, there is a genuine motivation for rule-based views to maintain practical equivalence independent, of course, of the costs of doing so.

Now I will explain how practical equivalence would be achieved/maintained on the part of rule contractarianism with respect to THE CONFLICT CASE. And from this, it will be apparent the costs to such a view in maintaining equivalence. Now, there are actually two ways in which rule contractarianism would arrive at the same result in THE CONFLICT CASE (and hence be practically equivalent to act contractarianism with respect to this case). The first way is this: A rule-based view would specify a single rule that got the right result in THE CONFLICT CASE – namely, the permissibility of (and perhaps the requirement of) the acting agent saving the child. I will argue that according to this method, by getting the result which aligns with our shared intuitions, it recommends a rule which violates THE MORAL RULE THESIS. Furthermore, this example shows that there are other cases in which a rule-based view under this recommendation will arrive at decisions which, although align with our shared intuitions, are based on rules which violate THE MORAL RULE THESIS. I will provide the details of this argument in the remainders of this section.

The second way for maintaining equivalence in this case is this: A rule-based view would specify numerous rules. None of these rules would violate THE MORAL RULES THESIS. As such, such a view could be practically equivalent (that is, arrive at the same results as act contractarianism) and yet be a genuine alternative to act contractarianism; the rules of such a view would be practical moral rules according to THE MORAL RULE THESIS. Now, of the multiplicity of these rules, there would be one which upon following it, one would do what aligns

with our shared intuitions. The problem, as I will illustrate, though, is this: In order to determine which rule is the one to follow, there must be some way to adjudicate between the multiplicity of moral rules. In other words, how are we to determine which rules are violable in THE CONFLICT CASE and which rule is such that one ought to follow it? I will argue that adjudicating amongst the multiplicity of rules makes a rule-based view – just so long as it maintains practical equivalence to an act-based view – become a view which is not genuinely a rule-based view at all. Now that I’ve explained the nature of my argument against rule-based views (in particular, rule contractarianism) which maintain practical equivalence to act-based views, I will now proceed with the details of the arguments.

Let’s consider in detail the first of the two recommendations in maintaining practical equivalence (with respect to THE CONFLICT CASE). Recall that the first recommendation is: (I) A rule based-view would specify a single rule that got the right result in this case – namely, permissibility of (and perhaps the requirement that) the agent saving the child. This single rule could take various forms. For example, it might look like this: “Whenever one is presented with the circumstances of having to break a promise to meet with a friend for something casual like a lunch, and also having to break a promise to meet a distressed friend, and also having to not help a slightly injured person one could easily help in order to save a child from drowning, one should/must save the child.” This rule contains 46 words with at least three letters. Clearly, this rule violates THE MORAL RULE THESIS. In THE CONFLICT CASE, in order for a rule-based view to maintain practical equivalence and get the correct result by specifying a single rule, it will have to specify a rule violating THE MORAL RULE THESIS. And so, such a view is not a genuinely rule-based view; it’s not that it is a bad view, it’s just that it’s not an alternative to my view.

Second recommendation in maintaining practical equivalence: (II) Instead of specifying a single rule, a rule-based view could specify numerous rules, such as: “Save innocent people from dying when doing so comes at little cost to oneself and to others”, “Do not break promises”,

“Help innocent injured persons when doing so comes at very little cost to oneself and others.”

Now, by rescuing the child, person A does not violate the first rule, but violates the other two.

It’s important to point out that a rule-based view where a person acts permissibly and yet in doing so violates one of its rules is not problematic for that reason; there is nothing problematic about a rule-based view which contains rules that will be violated in doing what is required/permitted by morality. Instead, the problem is this:<sup>53</sup> In any given case, what is the standard determining which rule is to be followed, and which may be violated? A rule-based view going down this second route must somehow be able to find a standard which adjudicates between the multiplicity of its rules.

Let me expand on this problem by discussing the alternatives for the rule-based theorist in answering to the problem of adjudicating between rules. Either rule contractarianism can take the position that the standard is itself an output of the contracting process, or it can take the position that it is not an output. If the latter, then rule contractarianism is an incomplete theory. The standard which determines which rules apply and which may be violated from case to case (when such violated rules still apply) would not be decided by the contracting agents; since moral evaluations are to be solely determined by the contracting agents according to contractarian views, this would make rule contractarianism incomplete. Now, if the former option is embraced by rule contractarianism, then there is the following problem: Either (i) the contracting agents determine what the standard is by looking at every single possible hypothetical case from a reasonably finite set (for example, a set of 10, 000 distinct cases, each different in a morally

---

<sup>53</sup> Actually, the problem is two-fold. In addition to the problem that I address in the body of the paper, there is another, very much related, problem. I state this related problem in a footnote for various reasons, one of them being is that I don’t think it is crucial in establishing my case against rule-based views. The problem is this: There is the question of how an agent would know that the rule to follow is “Save innocent people from dying when doing so comes at little cost to oneself and to others”, and not the others, such as “Do not break promises.” Perhaps the agent could guess, whether it be a random or an educated one, and if the agent guesses correctly, then there is no problem, as she does what is permissible (or, required). It would be rather spectacular, though, if all agents would always be able to successfully make guesses. Now, instead of pursuing this first problem further, I will grant to the rule-theorist that she will be able to explain according to her view how an agent would be able to know, in any given case, which rules are to be followed, and which rules may be violated in order to do what is morally permissible (or, required).

relevant way) in order to determine a standard that could be used in any given case in order to determine which rules are to be followed and which may be violated, or (ii) the contracting agents do not examine every possible case from the reasonably finite set, and instead are supplied enough information (for example, that pain is in general bad for people, and that killing other people results in bad consequences, and that lying and stealing lead to, in general bad results) in order to generate a reasonably adjudicable set of moral rules.

Let's consider (ii) first. I will suggest and motivate the idea that by following (ii), a rule contractarian account will not be able to maintain its practical equivalence. As such, by going this route, rule contractarianism will not suffer from problems of practical equivalence, as it will not be practically equivalent. Why won't such a rule contractarian account maintain practical equivalence? It is this: By not being presented with all possible cases from a reasonably finite set of cases (10,000 cases), it will arrive at a standard that prioritizes some rules over others in such a way that by following the prioritized rules, one will act in a way contrary to what is recommended by an act contractarian account; this then means that rule contractarianism would not be practically equivalent to act contractarianism. Nevertheless, it will be apparent in section 3.5 that by going down the route of (ii), rules will be arrived at in a way that makes rule contractarianism incoherent.

Let's now consider (i). If rule contractarianism follows this route, then I contend that, although it maintains practical equivalence, the standard for adjudicating rules will make rule contractarianism (or to generalize even further, rule-based views in general) not a genuine rule-based view. Why think this? Since the contracting agents will have considered all possible cases (from the set of 10,000 cases) in the search of a standard adjudicating between rules, they will also have generated rules that are so long and complicated that they violate THE MORAL RULE THESIS. Recall that in finding the adjudicating standard, the contracting agents are presented with each case, and thus they are able to formulate rules which determine the permissibility of actions with respect to each case. Yet these rules will be rather long and complicated.

### *3.4.3 Summary concerning practical equivalence*

Certain costs come with rule-based views maintaining practical equivalence with act-based views. More specifically, we saw that certain costs come with rule contractarianism maintaining practical equivalence with act contractarianism. Because of these costs, a rule contractarianism is motivated to alter her theory in such a way that avoids practical equivalence. In the following section, I will argue that rule contractarianism then becomes either incoherent or counterintuitive. In making clear what I mean by this charge of incoherence as well as why it applies to rule contractarianism, I will first examine the charge that rule utilitarianism is incoherent upon it not being practically equivalent to act utilitarianism.

## 3.5 INCOHERENCE (RULE-WORSHIP) AND COUNTERINTUITIVENESS

### *3.5.1 Introductory explanation of the charges*

I begin this subsection by exploring the rule versus act utilitarianism debate with respect to the second horn of the dilemma as mentioned at the beginning of section 3.4. Recall that the dilemma generalized to any rule-based view is this: Either a rule-based view rule avoids the charge of incoherence at the costs of being practically equivalent to act utilitarianism (where such costs were explained in section 3.4), or it avoids the charge of practical equivalence at the cost of being incoherent. Given that I've explained the costs of practical equivalence, it now makes sense to explore the costs of avoiding practical equivalence. As already mentioned, the costs of avoiding practical equivalence can be broken into three categories. First, practical equivalence can be avoided at the cost of being indeterminate. Second, a theory can avoid practical equivalence at the cost of arriving at results which do not align with our shared intuitions – that is, by arriving at implausible results. Third, a theory can avoid practical equivalence by being incoherent, where such a theory amounts to rule-worship. (Note that the first two charges can be explained as the theory being “counterintuitive”, in that it results in counterintuitive results, where either it is indeterminate, when, in fact, we think there is a determinate result, or where it results in a decision which runs counter to our shared intuitions.)

### 3.5.2 *The Negative Case*

In this subsection, I am going to present a case in which I evaluate rule contractarian views in light of how they would arrive at results concerning such a case. Such an evaluation will come in the next subsection. In this subsection, I merely present the case.

Consider the following scenario, which I will call THE NEGATIVE CASE: Imagine a large maze, perhaps 10 acres in area, where there are 10 dead-ends, each terminating in a small pond. Now suppose that you are aware of a child drowning in one of the ponds, and you are the only person who can save him. One problem, though, is that you don't know which pond he is drowning in, and you don't know the maze well, as you've never explored it. In order to get to the child in time to save him, you need a map to follow, and you also need to know which pond she is in. Suppose further that you know that the only way to get all of the required information is to perform these steps sequentially:

- (a) Since it is after open-hours, you have to lie to the first security guard at the entrance of the maze complex. You tell the guard that you left your coat in the lobby of the complex. She lets you through.
- (b) You then have to get past the second security guard by stealing her keys, which then lets you into the park surrounding the maze. You do this by slyly talking with her about how you lost your coat, and that how silly it was for you to bring a coat on a 55 degree day. By doing this, you are able to get close to the security guard, and slip the keys off of her.
- (c) Then once in the park, in order to get into the information building leading into the maze, you have to somehow get past the third security guard. Since the security guard will not be duped, and does not believe you that there is a child drowning, you have to assault him with your mace; you do this, and then lock him in a closet.
- (d) Finally, the way to physically get the map, and then the information about which the pond the drowning child is in, you have to torture the information officer. The information officer, although not responsible for the child being in the pond, does have

all of the required information. Note that he knows that the child is drowning. The information officer has been paid \$10,000 to keep this information from you.

I take it that it is intuitive that, minimally, it is permissible that you follow steps (a) – (d) in order to save the drowning child from dying. Now, I do understand that this is a difficult moral case, at least in the sense that various actions are required (for example, torturing) in order to save the child. But the difficulty of doing these things does not make the actions morally impermissible; at most, the difficulty may show that such actions are not morally required.

### *3.5.3 Incoherence (Rule-worship) or Counterintuitiveness*

In this subsection, I discuss the problems that rule contractarianism has in light of its departure from maintaining practical equivalence, and do this in light of THE NEGATIVE CASE. First, I will contend that there are three possibilities for what the rule-based contracting agents would do with respect to the implications for THE NEGATIVE CASE. Either (i) they would not come to an agreement at all or (ii) they would come to an agreement that would counter our shared intuitions about what the right thing to do is. The third possibility, (iii), is that they agree to results which run counter to our shared intuitions, but embrace such results because they are gotten by rules which they “worship”; it is this last possibility in which such a theory is incoherent (as opposed to implausible or counterintuitive). Before discussing these possibilities, I will discuss the kind of rules that a rule contractarian view would generate in light of THE NEGATIVE CASE.

In light of THE NEGATIVE CASE, we could reasonably imagine the following rules being agreed upon as amongst the list of correct rules: (1) Do not lie; (2) Do not steal; (3) Do not harm innocent persons unless in self-defense; (4) Do not torture others. And there would be a plentitude of others. Each of these rules could even have exception-clauses built into them (but of course, not to the point that they become so complicated that they violate THE MORAL RULE THESIS).

But now let's recall THE NEGATIVE CASE. Recall step (a) from the case. In following step (a), one is violating rule (1), and that in following step (b), one is violating rule (2), and that in following step (c), one is violating rule (3), and that in following step (d), one is violating rule (4). In order to do something that is intuitively morally permissible (and perhaps, morally required), one has to break the rules of "Do not lie", "Do not steal", "Do not harm innocent persons unless in self-defense", and "Do not torture others." What I've just said is true with respect to THE NEGATIVE CASE. I hereby suggest that whatever list of rules, there is a case in which doing the thing that is morally permissible (or perhaps morally required) violates all of the rules of the list.

So now let's consider possibility (i) in detail. Here, we would have contracting agents knowing what their position is with respect to the moral case at hand (i.e., there is no veil of ignorance). With respect to THE NEGATIVE CASE, there would be individuals knowing that they are one of the security guards or the information officer or the child drowning or the person doing the acting. Given that we are exploring the possibilities for rule contractarianism, what kinds of rules would they agree to? Suppose that they agree to a list of pro tanto rules, rules very similar, if not identical, to rules (1)-(4) mentioned above. Given that these rule are pro tanto, there is a further issue of which rules are to be followed and which are to be violated. Yet given that the contracting agents are not placed behind a veil, it is also plausible to think that no agreement would be made. For example, the security guards would not want to be lied to, or stolen from, or assaulted; they would agree to rules forbidding such actions. And the information officer would not want to be tortured, thus an acceptance of the rule against torturing. Yet the child would not agree to any of these rules, as agreement to such rules would make it permissible that she not be saved from dying.

In effect, according to possibility (i), we have a contractarian view endorsing a crude ordering of pro tanto rules, one where such ordering is so indeterminate that results are not achieved in moral cases where we would think that such results can be achieved. We think that in

THE CONFLICT CASE there is a determinate result (namely, that it is permissible (or in fact required) to save the drowning child). Note importantly that the charge here of indeterminacy does not require the endorsement of the idea that all moral cases are determinate. Instead, the charge is merely that in *this* case, there is intuitively a determinate result, and yet according to option (i), no such result is achieved. This makes rule contractarianism according to (i) counterintuitive/implausible.

Now let's consider (ii). Here, once again, we have the same contracting agents not placed behind a veil of ignorance. (Perhaps these contracting agents are more like Gauthier's contracting agents.) Here, we might have the security guards and the information officer pressuring the child into agreeing to rules (1)-(4), rules which would make it impermissible to save him (the child) from drowning; they might threaten him by certain force, by saying that he will be tortured by them before being drowned. It would then seem rational for the child to agree to rules (1)-(4), where agents following through with such rules would result in drowning as opposed to being tortured and then drowned. Here, we have a counterintuitive result – that it is permissible that the child allowed to be drowned. Basically, the charge against rule contractarianism according to option (ii) is that it recommends an action which runs counter to our shared intuitions.

Before moving onto option (iii), I should note some other possibilities, ones that are minor. Suppose that the contracting agents are, in fact, placed behind a veil of ignorance. In this case, they do not know whether they are a security guard, the information officer, the child, or the acting agent. Not knowing this, they may or may not agree on rules (1)-(4). Either they come to an agreement or not. If no agreement, then once again, we have the problem of indeterminacy in this case. If they do agree, this may be because they agree on rules (1)-(4) or some other set of rules. If they agree upon rules (1)-(4), which seems likely, then we run again into the problem of counterintuitiveness – that is, it will be impermissible to save the child because doing so will violate rules (1)-(4). If they agree upon some other set of rules, then my rhetorical question is

this: What would these rules be? They would have to be rules which allowed for the violation of rules (1)-(4) (for example rules against torturing) or they would have to be rules with so many exception clauses that they would violate THE MORAL RULE THESIS.

Finally, let's consider (iii). Note that the option of agreeing to rules which would amount to a result aligning with our shared intuitions is off the table. Such an option was already discussed with respect to practical equivalence. According to this option, we have a rule contractarian view embracing short and simple rules (which do not violate THE MORAL RULE THESIS), and yet in doing so, despite such endorsement leading to implausible results, such a rule contractarian view maintains such rules at the cost of rule-worship. The idea is this: Why maintain such a rule-based view if doing so leads to implausibility? If a view maintains its rule-based account despite implausibility, then it can be charged with rule-worship (or incoherence). To put things dramatically, imagine an account of rules in which any application of the rules would require that one not kill one person in order to save 1 million. Then such an account would be guilty of rule-worship/incoherence.

#### *3.5.4 Some concluding remarks*

I have now argued that rule contractarianism either violates practical constraints on what counts as practical moral rules or it runs into the problems of either (i) indeterminacy (in cases where we expect determinacy), (ii) implausibility/counterintuitiveness, or (iii) mere rule-worship. In light of these problems, the only alternative within contractarianism is to endorse an act-based view. Furthermore, Rawlsian ethical act contractarianism is appealing as an act-based view in virtue of its endorsement of the veil of ignorance. Without the veil, an act-based view (as I discussed at the beginning of this chapter) will be rather implausible, in that there would be numerous cases in which agreement would not be achieved. (Recall conflict cases in which some actions will make others worse; knowing that you will be worse off, there is little or no reason to agree to such an action, unless there is some alternative which makes you even worse off). But with a veil, the problem of no-agreement is solved. Not knowing one's position, the contracting agents will use

the principle of maximizing expected utility to make decisions and come to agreement with respect to actions.

## CHAPTER 4 – RAWLSIAN ETHICAL ACT CONTRACTARIANISM AND THE VEIL OF IGNORANCE

### 4.1 INTRODUCTION

In the previous chapter, I argued for my departure from the view that the objects of agreement of the contracting agents are rules/principles. Instead, the objects of agreement are actions and practices. Any way of construing Rawls' theory of justice as an ethical theory would be one in which the objects of agreement are rules/principles. As such, I made a significant departure from Rawls, as well as from most of the literature on contractarian ethical theories. In this chapter, I make further departures from Rawls' theory. In particular, I depart from the assumption that the contracting agents are mutually unconcerned. And although we can still think of the contracting agents as being self-interested, what's more important is to consider their preferences. Many people's preferences include the well-being of others, and to this end, even in thinking of the agents as self-interested, they will still make decisions which reflect their concern for others. I also depart from Rawls in terms of his endorsement of primary goods. I deny that what the contracting agents are trying to do is to make decisions which advance primary goods. Instead, they make decisions in light of what will maximize their own expected utility. In the end, though, I give a defense of a veil of ignorance in general. More so, I argue for the power of my account in light of the problems that have typically faced Rawls' theory in light of his employment of a veil.

Recall that in chapter 2, I argued that two major non-Rawlsian ethical contractarianism views were inadequate in light of three problems (namely, the Gangster Inclusion Problem, the Outlier Problem, and the Numbers Problem). In this chapter, I directly argue for a Rawlsian ethical contractarian view by (i) showing that it can give adequate solutions to each of the three problems and (ii) defending it against certain objections leveled against it. More specifically, the Rawlsian component I defend is the employment of a veil of ignorance. Recall that my characterization of my account is as follows:

Rawlsian ethical act contractarianism: An action A is morally permissible if and only if, from behind a veil of ignorance, A would be agreed to by all contracting agents each of which are the way they actually are in real life, except that they are perfect expected individual utility maximizers.

For the purposes of this chapter, I am defending the component of Rawlsian ethical act contractarianism which employs the veil of ignorance. Nevertheless, it will become more apparent how the principle of maximizing expected utility is a powerful and enticing principle for the contracting agents to use just so long as they are placed behind a veil of ignorance.

The methodology of this chapter is as follows. In section 4.2, I examine certain objections that have been made to Rawls' account. Where there is any credibility to these objections, I will take note. Given that these objections will have strength, it is only appropriate that my account deviates from Rawls' account. I will then offer a principled deviation from his account, where what I do is explain how a sophisticated version of the principle of maximizing expected utility in combination with the veil of ignorance is to be used by the contracting agents in arriving at moral decisions; in essence, I will argue that these two features, when combined, provide the fundamentals to a plausible ethical contractarianism. The principle of maximizing expected utility is first discussed in 4.3. For the remainder of 4.3, I reply to the earlier objections in terms of the appropriate departures I make from Rawls' view, highlighting how a combination of maximizing expected utility and the veil work to respond to these objections.

Recall that in Chapter 2, I discussed the shortcomings of alternative contractarian ethical theories in terms of three problems, namely the Gangster-Inclusion Problem, the Outlier Problem, and the Numbers Problem. In 4.4, I show how my account, primarily because of its employment of a veil of ignorance, is able to account for those three problems. In 4.5, I further motivate my account in numerous ways. First, I explore a certain problem that seems to be unique to an account like mine which has the contracting agents reason according to the principle of maximizing expected utility. Second, I motivate my account – both its veil component and its

maximizing expected utility component – in terms of its ability to get the right results concerning two kinds of moral cases. Third, I explain how my account does not rely on the appeal to analogous moral cases in order to arrive at decisions concerning moral permissibility. Fourth, and finally, I talk about how my account handles certain cases dealing with future generations.

Yet before moving onto all of, it is appropriate that I motivate my endorsement of employing a veil of ignorance. Suppose that I were asked the question: Why do you side with Rawls in employing a veil of ignorance? In what remains in this introductory section, I will motivate the veil of ignorance. And I will do this via an intuitive example. The main feature that motivates the veil of ignorance is that in employing it, one is able to achieve impartiality, where this is an essential component of any ethical theory.

As for an intuitive example of the power of an account employing a veil of ignorance, let's consider slavery. Slavery is a great wrong. Any adequate ethical theory must be able to give the result that the practice of slavery is morally wrong.<sup>54</sup>

Suppose that the contracting agents are deciding on whether the practice of slavery is permissible. The contracting agents would consider the various outcomes of individuals under a practice of slavery. From behind a veil of ignorance, a contracting agent would not know whether she would be a slave or a master (or, perhaps neither a slave nor a master) were there to be a practice of slavery. Although there are great benefits from being a master, there are also great costs to being a slave. If one thinks about this, then one will realize that many great facets of life are denied to a slave. The costs of being a slave are very great. As such, the contracting agents would not agree to a practice of slavery. They would agree that a practice of slavery is

---

<sup>54</sup> I want to emphasize that there may be certain conditions under which slavery is morally permissible. For example, suppose that a nation (or group) of people threaten that if they are attacked by another nation (or group), then if they successfully counter the attack and are able to conquer the attacking nation (or group), they will enslave them. Perhaps such slavery is morally permissible. Then again, there are a number of issues which arise. First, there is the issue of whether this would be genuine slavery, as the attacking group chose to attack upon being told that one consequence of attacking would be "slavery." A second issue has to do with whether the enslaving of dissenting citizens of the attacking nation is permissible. There are other issues as well.

impermissible/forbidden. And this is the result that aligns with our shared intuitions. As such, an ethical theory employing a veil of ignorance gains credit in virtue of getting the right result in such an important case. And it does this by achieving impartiality. In contractarian ethics, the main basis for showing that slavery is morally wrong is that by situating the contracting agents in a way that their decision is impartial, so too will their decision be the correct one in terms of morality.

## 4.2 PROBLEMS FOR RAWLS' VIEW

### *4.2.1 Objections arising from the relation between the veil and moral arbitrariness*

In this subsection I develop an objection based on substantive criticisms made by Gauthier, as well as Nozick, concerning Rawls' employment of a veil of ignorance. The objection arises because of the role that moral arbitrariness plays in motivating a veil of ignorance. I will clarify the notion of moral arbitrariness, in particular in relation to the worries that both Nozick and Gauthier have concerning Rawls' veil of ignorance.

As both Gauthier and Nozick remind us, Rawls employs a veil of ignorance to eliminate any features among the contracting agents that is arbitrary from a moral point of view (from here on, "morally arbitrary"). Recall the slavery example I discussed at the beginning of this chapter. Here, the fact that a master has enough power over slaves to do whatever he desires/decides is morally arbitrary. And the fact that a slave does not have enough power to resist the coercion of his master is morally arbitrary. It is morally arbitrary in the sense that just because these are facts does not mean that the morally right thing to do should be based on these facts; just because the master has these powers does not mean that it is morally permissible for him to exercise those powers. The veil of ignorance seems to be rather powerful in eliminating knowledge of those particular features of the contracting agents which are morally arbitrary, such that they are able to arrive at the correct moral decisions. Impartiality is achieved because of the veil's role in

eliminating morally arbitrary facts and features from weighing in on the decisions of the contracting agents.

Now, in order to understand objection before us, let us endorse a distinction Nozick makes with respect to moral arbitrariness. Nozick writes:

Here we see an ambiguity in saying that a fact is arbitrary from a moral point of view. It might mean that there is no moral reason why the fact ought to be that way, or it might mean that the fact's being that way is of no moral significance and has no moral consequences.<sup>55</sup>

Going back to the example of slavery, we can highlight the distinction. The fact that a person is a master over slaves is morally arbitrary in the first sense – in that there is no moral reason why the master ought to be a master. But the fact that a person is a master over slaves is not morally arbitrary in the second sense. The fact that someone is a master over other persons is of great moral significance, and has great moral consequences for both the slaves and the master.

Further context is needed now in order to understand the objection that Nozick has in mind. Nozick is more primarily concerned with the veil's exclusion of knowledge of individuals' natural endowments – those features of a person relevant to how well they may fare in the market, e.g., in being able to acquire and labor on natural resources. Nozick writes:

Why should knowledge of natural endowments be excluded from the original position? Presumably the underlying principle would be that if any particular features are arbitrary from a moral point of view, then persons in the original position should not know they possess them. But this would exclude their knowing anything about themselves, for each of their features (including rationality, the ability to make choices, having a life span of more than three days, having a memory, being able to communicate with other organisms like themselves) will be based upon the fact that the sperm and ovum which produced them contained particular genetic material. The physical fact that those particular gametes contained particular organized chemicals (the genes for people rather than for muskrats or trees) is arbitrary *from a moral point of view*; it is, from a moral point of view, an accident. Yet the persons in the original position are to know some of their attributes.<sup>56</sup>

The idea here is that *if* the grounds for excluding knowledge of individual natural endowments is that knowledge of any morally arbitrary feature is to be excluded (as natural endowments are

---

<sup>55</sup> Robert Nozick. *Anarchy, State, and Utopia*. (1974). p. 227.

<sup>56</sup> Nozick, p. 227.

morally arbitrary features), *then* this will be problematic, as the contracting agents are to know at least some features about themselves, features which are nevertheless morally arbitrary.

At this point in the dialectic, Nozick points to the distinction between the two senses of moral arbitrariness. If Rawls wants to allow for the contracting agents to know some of their morally arbitrary features in the first sense (e.g., that they are rational and that they are able to communicate with one another) on the grounds that they are not morally arbitrary in the second sense (that such features have moral consequences), then he will have to allow for them to have knowledge of their natural endowments. (Natural endowments are not morally arbitrary in the second sense.) And this is something Rawls does not want to do for the sake of his egalitarian theory of justice. So, the objection is a dilemma: Either Rawls' veil will exclude too many relevant features or, in including relevant features, it will also have to include knowledge of one's own natural endowments.

The objection against Rawls' veil can be further amplified by what Gauthier says about it. Gauthier writes:

For in his argument the epistemic effect of ignorance of one's identity has an ontological significance quite incompatible with conceiving persons as actors. A person's identity is in all respects a contingent matter. But this contingency is not morally arbitrary, for morality is and can be found only in the interaction of real persons individuated by their capacities, attitudes, and preferences. In our analysis we take the individuality of persons seriously.<sup>57</sup>

Here we have an appeal to Nozick's second sense of moral arbitrariness. Although a person's capacities and preferences are morally arbitrary in the first sense, they are nevertheless not arbitrary in the second sense. Because these kinds of features of persons and their relevance to how persons interact with each other are what morality is all about, the contracting agents should have knowledge of their own capacities and preferences.

Notice, though, that Gauthier's objection is distinct from Nozick's. Nozick points out that if Rawls allows knowledge of features not morally arbitrary in the second sense (but morally

---

<sup>57</sup> Gauthier. *Morals by Agreement*. p. 257.

arbitrary in the first sense), then he will also have to allow knowledge of natural endowments (as they are not morally arbitrary in the second sense). Gauthier objects that Rawls doesn't even allow for knowledge of features which are morally arbitrary in the first sense but not in the second sense.

In light of Nozick and Gauthier, the objection to Rawls' employment of the veil of ignorance is rather compounded. We have Nozick's dilemma. And we have Gauthier's point that the decisions of the contracting agents must be made in light of knowledge of their own capacities and preferences, since the stuff of morality just is the stuff of persons interacting with each other in terms of their differential capacities and preferences. And according to Rawls' veil, knowledge of these differential capacities and preferences is not known by the contracting agents.

#### *4.2.2 Nagel's objection and Hart's objection*

I now turn to two very-much related objections to the veil of ignorance, one presented by H.L.A. Hart and the other by Thomas Nagel. Let's start with Nagel's objection. Nagel objects that the veil of ignorance endorsed by Rawls is so thick that it rules out some morally relevant information to the contracting parties – namely, it rules out some conceptions of the good that are not served by the mere primary goods.<sup>58</sup> Primary goods are, according to Rawls, “things that every rational man is presumed to want.”<sup>59</sup> This means that the primary goods “have a use whatever a person's rational plan of life.”<sup>60</sup> Finally, as for a list, “the chief primary goods at the disposition of society are rights and liberties, powers and opportunities, income and wealth.”<sup>61</sup> There are other primary goods, such as health and intelligence, but Rawls does not think that these are as strongly influenced by social structures as the ones previously mentioned, such as the possession of rights and social powers. The concept of primary goods is essential to Rawls' theory, since they influence the choice of agents from behind a veil of ignorance. The contracting

---

<sup>58</sup> Nagel's “Rawls on Justice” in Daniels 1975, pp. 1-16.

<sup>59</sup> Rawls 1971, p. 62.

<sup>60</sup> *Ibid.*, p. 62.

<sup>61</sup> *Ibid.*, p. 62.

agents, although ignorant of whatever position they may occupy in society, will nevertheless know what the primary goods are, and subsequently will make decisions in the prospect of securing those goods. To put it this way, the contracting agents will make decisions to make sure that they have rights, that they have powers and opportunities, and that they have an acceptable amount of income and wealth.<sup>62</sup>

Nagel's objection amounts to something like this: Requiring that the contracting agents make decisions with mere knowledge of primary goods excludes the morally relevant information of the conceptions of what's good by actual, particular people. Furthermore, a focus on securing primary goods, as listed above, can lead to two things: (1) The primary goods will serve some persons more than others in fulfilling particular conceptions of the good. (2) The primary goods may well counter particular conceptions of the good. To highlight these points, consider the following passage by Nagel:

[the primary goods] are less useful in implementing views that hold a good life to be readily achievable only in certain well-defined types of social structure, or only in a society that works concertedly for the realization of higher human capacities and the suppression of baser ones, or only given certain types of economic relations among men. The model contains a strong individualistic bias, which is further strengthened by the motivational assumptions of mutual disinterest and absence of envy. These assumptions have the effect of discounting the claims of the conceptions of the good that depend heavily on the relation between one's own position and that of others...<sup>63</sup>

Notice that the last comment quoted supports a view like (2), from above, where some conceptions of the good may even be ruled out by the advancement of the primary goods.

Indeed, Nagel is right that caste societies, as well as ones based on perfectionist aims, will be ruled out by contracting agents reasoning making decisions in light of knowledge of the (i.e., Rawls') primary goods.<sup>64</sup> This can be no complaint, though, as intuitively we think that these are unjust societies.

---

<sup>62</sup> Note that I will later discuss the inherent ambiguities in Rawls' position on this.

<sup>63</sup> Nagel's "Rawls on Justice" in Daniels, p. 9.

<sup>64</sup> I should be a little more nuanced with my claim here. According to my account, some caste societies will, in fact, be unjust, but this would only because of the desires of the individuals comprising such

In relation to (2), Nagel's objection has more force, though, if we construe it another way. The primary goods will exclude the permissibility of those living lives according to caste arrangements or according to perfectionist aims, such as leading a life of chastity. If someone consents to her caste position, or if someone chooses to lead a chaste life, certainly it would be illegitimate to initially conclude that these people are doing something morally wrong. So, if Rawls' primary goods account excludes these possibilities – by deeming them to be morally impermissible, or unjust – then his veil of ignorance would exclude morally relevant information in an objectionable way. The move to make is to depart from Rawls in terms of primary goods. From behind a veil of ignorance, the contracting agents must be given sufficient enough information about particular conceptions of what's good instead of relying on a primary goods account.

Before turning to Hart's objection, it's important that we focus on (1) from above, where the effect of focusing on primary goods is that some are served better than others in advancing particular conceptions of the good. It's important because perhaps Nagel intends (1) more than (2). As such, perhaps his more fundamental objection centers around (1). Let's consider the primary good of wealth. Certainly securing more wealth is important for many people. But for others, those trying to perfect themselves under vows of poverty for example, such a primary good will do very little good. It will do nothing to advance their own particular conception of what's good. And it seems now unfair that from behind a veil of ignorance, the contracting agents are advancing some goods which will advantage some persons over others. Once again, we should abandon a primary goods account. And once again, this leaves us with a challenge. The contracting agents while behind the veil of ignorance must be able to be given sufficient enough information which will fairly advance each person's conception of the good.

---

societies – more specifically, that where desires are satisfied at the expense of other desires being frustrated, the contracting agents would agree to this.

Now let's consider Hart's objection.<sup>65</sup> Although Hart actually offers a number of criticisms, the one that I will examine concerns more specifically the priority rule of "sacrificing liberty only for the sake of liberty." What does it mean to sacrifice liberty for the sake of liberty? One example provided by Rawls, and mentioned by Hart, is that of sacrificing the liberty to speak whenever we like at a debate so that we have the liberty of effectively debating, since a debate where participants spoke whenever they wanted to would be uncontrolled, chaotic, and fall apart. Here, there are two conflicting liberties: liberty to speak whenever we like at a debate and the liberty to be member to an effective debate. It seems clear which liberty is to be sacrificed for the sake of the other.

Here, Hart rephrases the choice between the two liberties in Rawlsian terms:

To put the same exceedingly simple point in the 'maximin' terms which Rawls often illuminatingly uses, the worst position under the rule (being restrained from interruption but given time to speak free from interruption) is better than the worst position without the rule (being constantly exposed to interruption though free to interrupt).<sup>66</sup>

To summarize, "the representative equal citizen" in the original position would rationally prefer the worst position under the rule of being restrained over the worst position without the rule of being restrained. (Recall from Chapter 1 the definition of "the representative equal citizen.") But then Hart has us consider cases where there are differences in what real-world agents prefer. For example, "some people may prefer freedom of movement not to be limited by the rights of landowners supported by laws about trespass; others, whether they are landowners or not, may prefer that there be some limitations."<sup>67</sup>

Perhaps the parties would be able to settle the conflict in preference by basing their decision on what the representative equal citizen would rationally prefer. Yet rational preference

---

<sup>65</sup> In the preface for the revised edition of *A Theory of Justice*, Rawls writes that "one of the most serious weaknesses <in the original edition> was in the account of liberty, the defects of which were pointed out by H.L.A. Hart in his critical discussion of 1973. Beginning with <Section> 11, I made revisions to clear up several of the difficulties Hart noted. It must be said, however, that the account in the revised text, although considerably improved, is still not fully satisfactory."

<sup>66</sup> Hart's "Rawls on Liberty and its Priority" in Daniels 1975, p. 242.

<sup>67</sup> *Ibid.*, p. 243.

of the representative equal citizen (i.e., of the Rawlsian contracting agent) is determined by the list of primary goods (that she is trying to get the most of) and any priority rule that she has knowledge for prioritizing goods when they conflict. In this case, the list of primary goods does not settle the conflict between preferences, and there is no priority rule to appeal to either. Both liberties (the liberties at conflict) are neutral respect to whether primary goods are advanced by them, or the other way around – whether the primary goods advance one liberty over the other. So, no such preference is had by the representative equal citizen. As such, the parties, who decide according to the preferences of the representative equal citizen, are incapable of arriving at an agreement. Just so long as Rawls endorses a primary goods account, it appears that there is no way around Hart's objection. And yet once again, though, I've said that we are to abandon the primary goods account. And to remind the reader of the challenge of abandoning primary goods: The contracting agents, just so long as they are placed behind a veil of ignorance, must be given sufficient enough information that will allow for them to make agreements which fairly allows for each person to pursue her own particular conception of what's good.

#### *4.2.3 Hare's objection*

In this section, I present a significant objection, raised by Hare, to Rawls' thoughts concerning who is and is not included in the contracting process – i.e., who gets included as a contracting agent. We can call the problem the Exclusion Problem, and it either can be considered to be a specific aspect of the Outlier Problem, or it can be considered a closely related problem. We can characterize the Exclusion Problem, in general, as follows: Because not all persons (and others who have moral standing) will be included in the contracting process, some persons (and others with moral standing) will not be properly represented in the contracting process; furthermore, the contracting agents will not be restricted from deciding in ways that counter our shared intuitions about how those excluded from the contract will be treated.

Now we will start with fleshing out Hare's objection. First, note that anyone introducing a contractarian ethical theory should inevitably say something about who is and is not included in

the contracting process. The group of those included in the contracting process – i.e., the contracting agents – are referred to by Hare as “the committee.” Since we will be discussing Hare’s objection, I will frequently refer to the contracting agents as “the committee.” Now, in introducing a contractarian account, one then should say who is included as members to the committee; in doing so, one is advancing “membership-restrictions”, as Hare puts it.

To understand Hare’s objection, we should also consider the distinction between possible people and actual people. Actual people include those who do in fact exist presently, as well as those who will definitely exist in the future, and as well as those who have existed. Possible people are those who could exist in the future, given that certain conditions are satisfied (and basically, the main condition is an act of conception of egg and sperm).<sup>68</sup> Hare does not really spell out the distinction in this way, or any way at all. But I think that it should be helpful to the reader to give some kind of reasonable characterization of the distinction.

Now, Hare cites passages supporting Rawls’ membership-restriction where (in Hare’s words), “Only people who actually do or will exist are allowed on the committee.”<sup>69</sup> Hare, I think, is right in ascribing this kind of restriction position to Rawls. Thus far, Hare hasn’t shown the damage Rawls has done to his account; it’s to this we turn.

The Exclusion Passage: Hare writes: This means that in Rawls’ system the interests of possible people are simply not going to be taken account of. This would seem to be crucial for questions about population policy and abortion, for example. The person that the foetus would turn into if not aborted, and the people who will be born if contraception is not practised, get no say if they are not actual people – i.e., if it is actually the case that

---

<sup>68</sup> I should note this: The notion of possible person Hare has in mind is narrower than other senses of possible persons that we could have in mind. For example, we could imagine a possible person that is conceived without the egg and sperm conception. More accurately and specifically, for example, we could imagine a possible world in which human beings as persons were conceived by some annual ritual in which gatherings sacrificed another human being in order to make the conception (i.e., creation) of another human being. So, according to this sense of possibility, persons are created in a way that doesn’t require egg and sperm conjunction.

<sup>69</sup> Hare, p. 99.

abortion and contraception are practiced. This would seem to have the curious consequence that, simply by performing an abortion, I can make sure that my act does not contravene Rawlsian justice, because I shall have not disenfranchised the abortee.<sup>70</sup>

The idea here is something like this: Because it is only those who do or will exist who have moral standing, according to Rawls, by performing an abortion we would be able to exclude those aborted fetuses from the committee, since by aborting them, they end up not existing as persons; we've ensured the fact that they will not exist, and hence are not actual people (in the future).

There are actually two different cases under consideration here. The first kind of case is one in which two people having sex with contraception do not then successfully "make" a fertilized egg, which then would develop into a fetus, and then into a born human being. This act of having sex with contraception then makes it the case that the possible person (designated by a specific sperm-egg combination) is not brought into existence as an actual person, and thus is excluded from membership on the committee (or even consideration by the committee).

Let's now consider the second case. In appealing to abortion, I think that Hare has in mind a category of fetuses in which their being aborted is thought by many of us to be morally impermissible. The category I have in mind is third-trimester fetuses. So, we are to consider cases of third trimester fetuses, and whether it is permissible to abort them when there is no threat to the pregnant woman, or where no great burden will be placed on her in merely giving birth to the fetus. I suspect many of us would consider it wrong to abort in such cases (although of course, whether or not it is in fact wrong is still disputable and up for debate). Recall that according to Rawls, only actual people (people who do presently exist and will exist) are to be given moral consideration, not mere possible people. Hare's argument then needs to proceed like this: Since third trimester fetuses are merely possible people, and since according to Rawls we are not to give moral consideration to merely possible people, it would automatically be

---

<sup>70</sup> Ibid., pp. 99-100.

permissible to abort them, all other things equal. Because this goes against our shared intuitions, there is something wrong with Rawls' account.

A final kind of objection can be raised in light of Hare's comments. This objection is based on a problem dealing with future generations. This is a question of whether Rawls excludes members of future generations in a way that biases them. If it's the case that future generations will be excluded from the contracting process in a way that they are biased, then this indeed will be a problem. We could imagine the contracting agents, knowing that they are not a member of future generations, agreeing upon policies, practices, and actions which would benefit themselves but place costs on future generations. If future generations are excluded in the way just discussed, then this indeed would be a great objection to Rawls' account.

#### 4.3 HANDLING THE OBJECTIONS LEVELED AGAINST RAWLS' ACCOUNT AND MAINTAINING A VEIL OF IGNORANCE

In the previous section, I discussed various problems with Rawls' account in his employment of the veil of ignorance. I do not deny that there are problems with his account, but I do uphold the idea that a correct contractarian view employs a veil of ignorance. As such, this section works to show that the veil of ignorance, when properly understood, is the correct way of constructing a contractarian ethical theory. Recall the objection constructed from Nozick and Gauthier's work. According to Rawls' notion of moral arbitrariness, (i) either Rawls' veil will exclude too many relevant features or, in including relevant features, it will also have to include knowledge of one's own natural endowments; (ii) Gauthier observes that the decisions of the contracting agents must be made in light of knowledge of their own capacities and preferences, since the stuff of morality just is the stuff of persons interacting with each other in terms of their differential capacities and preferences, and yet according to Rawls' veil, knowledge of these differential capacities and preferences is not known by the contracting agents. And recall that Nagel's objection and Hart's objection requires a rejection of primary goods as the thing that the contracting agents are

primarily trying to advance while behind the veil; the challenge is to show that the contracting agents while behind a veil are capable of making decisions without using primary goods as what they are advancing. Finally, we have Hare's objection. Fundamentally, it is an objection concerning the exclusion of certain persons from the contracting procedure, and more importantly, from the protection/consideration of the contracting procedure. As I explained, Hare's objection does not have the initial force that it advances. Those who, in fact, have moral standing are given consideration by the contracting procedure. What's problematic according to the objection is the fact that future generations, in general, are not accounted for by Rawls's account. The question is: How, behind a veil of ignorance, can the contracting agents give appropriate consideration to persons who will definitely exist in the future (but who do not exist as of yet)?

#### *4.3.1 Employing the Principle of Maximizing Expected Utility*

The rational choice principle that I endorse for the contracting agents is the principle of maximizing expected utility (PME). One apparently unattractive feature of the rational choice principle of PME is that on certain views, utility is reduced to simply states of pleasure and pain. One might think that this is a crude way to understand utility. I agree. But the kind of view of utility that I have in mind understands utility in terms of preference-satisfaction. This kind of way of understanding utility can be very rich in terms of the psychology of individuals. Now, some might think that preference-satisfaction will ultimately reduce to the pain and pleasure relativized to individuals, according to whatever it is that gives such individuals pleasure and pain. Whether such a reduction is correct or not is irrelevant. What's important is that there are a number of complex psychological features which are accounted for by a preference-satisfaction view that I have in mind. The psychology of human beings is rather complex, and one can only suspect that the psychology of others affected by moral actions is also rather complex; here I have

in mind members of an intelligent alien species.<sup>71</sup> Given that a reasonable ethical theory considers the permissibility of actions in terms of the consequences of actions, and given that such consequences can be spelled-out in terms of the effects on the complex psychology of moral patients, it only makes sense to incorporate the richness of psychological features in the assignment of utility, where such utility is essential to an ethical theory.

Now I will be more concrete. A preference-satisfaction view of utility-assignment like mine accounts for things like resentment upon being treated in certain ways, in addition to any kind of pain that might be experienced. It can capture desires that some persons may have concerning fairness. More specifically, many of us have a conception of moral rights. Although these conceptions may be misled in some way or another, there is still a sense in which the experience of rights violations (according to such conceptions) is a way of having one's desires frustrated. As such, my way of understanding utility accounts for rights violations, in that whenever person A's rights are violated according to the conceptions of B, this would amount as a frustration of B's desires (just so long as B cared that persons' rights were not violated). Many of us also have beliefs concerning moral desert. And so, whenever C is treated in a way that D believes C does not deserve to be treated, we can count this as a frustration of D's desire for desert to be satisfied. Now, because both notions of rights and moral desert have been discussed, and because such considerations vary from person to person as more important to some rather than others, I will generalize and just refer to the notion of fairness to include both the conceptions of fairness and moral desert. Overall, how these psychological factors determine utility assignments will become more apparent as I discuss specific cases (which are relevant in giving adequate replies to the objections to my veil-of-ignorance account).

---

<sup>71</sup> Note that I'm not saying that non-human animals are not affected by moral actions; in fact, they are moral patients as they have moral standing. But human beings are also moral patients, and, given the complex psychology of human beings, it only makes sense to determine the permissibility of actions in terms of the complex psychology that is affected by actions.

Note that in order for the contracting agents to consider people's resentments, and more specifically, the desires of many people that rights and desert-claims are respected is something that entails that the contracting agents are mutually concerned *in some sense*. Now, because the contracting agents are behind a veil of ignorance, there is a sense in which they make decisions unconcerned for others. But the utility assignments that they consider are ones in which mutual concern plays a significant role. Real-world agents are ones that care very much about how they do in relation to others, for one thing; hence, the notion of resentment is essential to utility assignment. And yet also, real-world agents (at least many of us) care that *others* are treated fairly (according to our conception of fairness). And as I've said, when others are not treated in a fair way according to our conception of others, this amounts to the frustration of desires for fairness, and such desire-frustration is morally relevant. Hence, negative utility is assigned to this. So, there is a sense in which the contracting agents are, in fact, mutually concerned, in that they concern the utility of real-world agents which is determined by their desires concerning how others prosper, both in relation to themselves and independently of themselves.<sup>72</sup>

#### 4.3.2 *Reply to Nozick and Gauthier*

In this subsection, I counter Nozick's and Gauthier's objection leveled against Rawls' veil of ignorance. I will do this in two ways. First, I will clarify just exactly what kind of information is provided to the contracting agents while behind the veil. Second, I will employ my preference-satisfaction utility view in order to show that other kinds of concerns are addressed properly.

To start, I will make a distinction between indexical and non-indexical knowledge allowed behind a veil of ignorance. Indexical knowledge takes this form: Contracting agent, A, knows that I, as A, am A, and A has capacities, x, and preferences, y. Non-indexical knowledge takes this form: Contracting agent A knows that A has capacities, x, and preferences, y, but does not know that as A, she is the one who is A. Now, all contracting agents are given maximal non-

---

<sup>72</sup> For different reasons for departing from a contractarian view which endorsing mutual unconcern, consider Peter Vallentyne's "Contractarianism and the Assumption of Mutual Unconcern" in *Contractarianism and Rational Choice*. (ed.) Peter Vallentyne.

indexical knowledge. They know the preferences of all contracting agents as real-world agents. Suppose that there are 100 real-world agents. Then there are 100 contracting agents each knowing of the capacities and preferences of each of the 100 real-world agents. More specifically, given any action involving any of the real-world agents, all 100 know about the utility-assignments of the possibly affected agents. What the contracting agents do not know is the indexical knowledge – i.e., they do not know whether they are the ones affected in the specific moral case presented to them.

How does this give a sufficient reply to the objection constructed from Nozick and Gauthier? Answer: The contracting agents are able to take into account how various agents are affected by various actions, and this accounts for how capacities may be taken advantage of, as well as how preferences are satisfied or frustrated. Furthermore, each contracting agent knows of the various features possessed by real-life agents that are, in fact, morally relevant. That is, whether such agents are rational, are sentient, whether they will live more than three-days, etc. Such facts are taken into account via their inclusion as non-indexical knowledge, but just not as indexical knowledge.

Now, the second way to reply to this objection is not fully distinct from the first. This should be apparent as I give this second reply. Second reply: There is a worry with both Nozick and Rawls that by employing a veil of ignorance, certain people will be taken advantage of by others in a parasitic way.<sup>73</sup> Once again, the worry is due to the fact that certain morally arbitrary features, in one sense, which are certainly not arbitrary in the second sense – in that such features have moral consequences – will be discarded in the reasoning of the contracting agents. As such, the results of the deliberations of the contracting agents will result in a way that biases some individuals over others. But here's the reply, in general. Once we see how the contracting agents are able to come to decisions with respect to specific cases, and with respect to their employment

---

<sup>73</sup> Certainly, if we are to employ the principle of maximin, then some individuals at the bottom will be advantage at the expense of others. Nevertheless, I have decided to reject such a principle as the correct rational choice principle for contracting agents while behind the veil.

of the principle of maximizing expected utility, we will see that satisfactory results occur – namely, results which cohere with our shared moral intuitions.

So, let's take an example. Here, I will take a two-person case. Suppose that person A has acquired 10 carrots for the month. Suppose that person B has acquired 2 carrots for the month. Now, both A and B both prefer as many carrots as possible. Nevertheless, the needed amount of carrots per month is just 1. Suppose the contracting agents are considering the action of B stealing 4 carrots from A. If B were to successfully do this, then A would have 6 carrots and B would have 6 carrots. Is B's action permissible? I stipulate this: A thinks this: B already has more carrots than is needed in one month (2 opposed to 1). Also, I (A) have worked hard for my carrots, even if it is the case that B worked hard for hers. Furthermore, yes, with me having 6 opposed to 10 carrots I can assign 6 opposed to 10 units of well-being to this happening to me. But I also will assign some more points of disutility given the fact that I have been stolen from, and that the person who stole from me already had more than what she needed; so, I will assign 2 points of negative utility points. B, on the other hand, assigns 6 utility points to stealing the carrots. But B, unless B is a sociopath, does not assign any further utility points to doing this; she does not get any further satisfaction from the fact of stealing from A. As such, the results are as follows: Stealing from A: A gets 4 points, whereas B gets 6 points. Not stealing from A: A gets 10 points, whereas B gets 2 points. According to PME, the contracting agents choose "not stealing from A", as this maximizes expected utility. Note that according to maximin, the contracting agents would choose "stealing from A", and yet this is counter to our shared intuitions.

Now, what does this have to do with Nozick and Gauthier? Well, it's this: We can make sense of how those who are taken advantage of by those "at the bottom" have done something morally wrong. Just because a person is doing not as well as another person does not mean that they are entitled (morally) to be advantaged by those doing better. Of course, sometimes this will be morally required of those on the top. But, with my sophisticated utility account (where

psychology of individuals is taken into account), we can get the intuitive results, where such results should align with the intuitions that Nozick and Gauthier appeal to in making their cases. In conclusion, there is nothing suspect to the veil of ignorance, but instead only to the rational choice principle of maximin that Rawls invokes. In addition, there is the possible problem of what knowledge is allowed while behind the veil, and yet I clarified this via the indexical vs. non-indexical distinction.

#### *4.3.3 Reply to Hart and Nagel*

My reply to this objection amounts to a rejection of Rawls' notion of primary goods. (Once again, this is a place where I advise the reader to recall the characterization of Rawls' primary goods that I discuss in Chapter 1.) Without the notion of primary goods, there just won't be a conflict of rational preferences between the contracting agents. Instead, they will merely know of the preferences of each individual involved in the case that they are examining and evaluating (and which they are to come to a decision with respect to). Let me be a little more concrete.

Suppose that we have the following case: There are three people involved. One of the people likes to be hit really hard by two other people and then injected with heroin from them. The other two people don't like the idea of using violence in this way. The contracting agents would then consider the supposedly differing preferences between the three. If the costs to the two who are candidates for imposing the beating and injecting of heroin is high enough, then it will turn out that it will be permissible that they do not do this beating and administration of drugs to the one wanting this. I take this to be an intuitive result. In general, my reply to the objection is this: There is no such thing as a conflict in rational preference while behind a veil of ignorance; whatever alleged conflict there is always gets settled by the application of PME.

My reply, once again, is simple. It replicates my reply to Hart's objection. It's that I depart from Rawls with respect to a primary goods account. If it is the case that persons under, for example, a caste society find value in being placed in a subservient role, then this is accounted for by the account that I give (with PME). The idea is this: A considerable amount of utility is

assigned to those individuals who prefer to be subjected to a caste system. That is, if they prefer this, then they are assigned a considerable/appropriate amount of utility to the position that they prefer. Now, as contracting agents are to reason in such a way that they are to consider the possibility that they are one of such members of a caste society, where in such a society a strong majority of those at “the bottom” nevertheless find a great amount of value in their position, I think that the contracting agents would agree to the arrangement of such a caste society. I think that this is the intuitive result that Nagel appeals to, and as such, the reply I give to his objection is, I think, correct. Now, many of us will reject this result – that it is permissible that there be such caste society practices. But such a rejection is based, I believe, in a certain mind-set that is shared by those who participate in societies who do not (explicitly) appeal to caste-arrangement ideas. The idea is this: If you’re in a caste arrangement society and you’re “cool with this” even if you are at the bottom of the arrangement, then this should be respected. With this said, I have given an adequate reply to Nagel’s objection to an employment of the veil of ignorance. Note that Nagel’s objection, I think, still stands against a veil of ignorance account employing Rawls’ notion of primary goods. But given that I depart from Rawls on the employment of a notion of primary goods, I think that I have adequately defended a veil of ignorance account (which does not appeal to primary goods) to Nagel’s objection.

#### *4.3.4 Reply to Hare*

Recall that the objection Hare raises is the Exclusion Problem. The idea is that by Rawls only allowing actual people – i.e., people who do exist or who will exist – to be represented in the contracting process, there will be a problem of excluding certain beings who should be given moral consideration. More specifically, there were the two cases that Hare mentions. The first is the possible person who would exist were it the case that contraception is not used. The second case is one of aborted fetuses. In both cases, it appears that if these beings are not actual people, then by aborting them or by using contraception, we can then exclude them from consideration of

the contract. Let me first explain, in general, the conclusions I come to with respect to Hare's objection.

I am now ready to respond, on behalf of Rawls, to the objection arising from the first case. My reply: To assume that any wrong being done here by having sex with contraception would be to assume that possible people have rights to be brought into existence as actual persons. This assumption just does not carry any weight in terms of our shared intuitions. Hence, at least in this case, Hare's objection has no force. As I will discuss later, the idea that possible people have moral consideration can be confusing. We have to make a distinction between considering whether a possible person has a right to be brought into existence as an actual person (which I claim most of us think is not the case) and considering whether upon bringing a possible person into existence as an actual person is permissible given the kinds of things that may be owed to them upon being brought into existence.

Now let's consider the second case again. The question is: Does Rawls' view commit him to the idea that by aborting fetuses (in particular, third-trimester fetuses) they are not to be given moral consideration? My answer is no. And this is because Hare has mistakenly forgotten that third trimester fetuses, for many of us, are already actual people. To be more specific, many of us believe that aborting third trimester fetuses is wrong *because* they are actual people (or more minimally, have strong moral standing), and there is no other outweighing reason – such as health of the pregnant woman or excessive burden on the woman – to judge otherwise. So, Hare's argument only works if we consider third trimester fetuses to be merely possible people. But there is good reason to believe that they are actual people. Since they are actual people, according to Rawls' account, it is not so easy to see how his contracting agents would decide that aborting them is morally permissible. In fact, with all other things equal, I suspect that Rawls' account could easily conclude that the abortion of third trimester fetuses is morally impermissible. By being placed behind a veil of ignorance, contracting agents would not know

whether they are a third trimester fetus or not. Acknowledging this possibility, the contracting agents would not risk being killed via abortion, as a being with significant moral standing.

From this discussion, we can see that a veil of ignorance is rather powerful in dealing with exclusion problems, as discussed above. While behind a veil, one does not know whether one is representing a being with moral standing that nevertheless would be excluded from other contractarian accounts. (Here, recall the Outlier Problem.) As such, one would not risk jeopardizing the interests/desires of the one she could be representing. If, for example, third-trimester fetuses have moral standing, then the contracting agents do not know whether they represent the interests of such fetuses; as such, they are not excluded. This can also be applied to non-human animals (as already mentioned with respect to the Outlier Problem.)

Finally, we must consider what Rawls has to say with respect to future generations. If future generations are not given the proper moral consideration according to his account, then this indeed would be a problem. Now, it is explicit, according to Rawls, that the contracting agents do know that they are all of the same generation. And he takes this further as follows:

Those in the original position know, then, that they are contemporaries, so unless they care at least for their immediate successors, there is no reason for them to agree to undertake any saving whatever. To be sure, they do not know to which generation they belong, but this does not matter.<sup>74</sup>

Why is this important? It's because that as Rawls presents it, the concern for future generations becomes subsidiary, at the least. If the contracting agents know that they are of the same generation, and furthermore know that they are considering issues relevant to their own generation or of future generations, then they may very well be biased against such issues concerning future generations. This indeed is a problem for Rawls' view.

Rawls does, in fact, offer a solution to this problem, but as I will readily claim, his suggestion does not even come close to anything remotely plausible. Rawls' suggests that in characterizing the contracting agents, we should "adjust the motivation condition. The parties are

---

<sup>74</sup> Rawls, p. 292.

regarded as representing family lines, say, with ties of sentiment between future generations.”<sup>75</sup>

First of all, this is unsatisfactory for the reason that I do not want the contracting agents to be construed as parties, or as representative citizens, or as family-lines. The contracting agents are individuals. Secondly, as contracting agents, some will be fathers and mothers, etc. But others will not be parents (at any point in their lives). Others will be fathers and mothers lacking the appropriate motivation condition of care for their offspring. These kinds of contingencies should be taken into account when constructing an ethical contractarian theory. The relevant question is: What would the contracting agents agree to with respect to their treatment of future generations when they do not know whether they are a parent who cares, or a parent who doesn't care, or who will never be a parent? If they know they are of the same generation, counterintuitive results may come about. As such, I think that a plausible construal of the original position will either (i) include those who are of future generations, or (ii) include trustees for future generations (namely, one will not know whether she is a member of the present generation, or whether she is representing someone of a future generation).

#### 4.4 HANDLING THE THREE PROBLEMS

In chapter 2, I had discussed the outlier problem, the gangster-inclusion problem, and the numbers problem. In this subsection, I will briefly show how a contractarian account like mine, employing a veil of ignorance, can provide natural solutions to each of these problems.

##### *4.4.1 A solution to the outlier problem*

Recall that I have characterized my contracting agents as perfect individual expected utility maximizers. This might seem to eliminate the inclusion of SLRCs (sufficiently low rational capacities) and non-human animals (and perhaps children) from considerations of moral matters. Additionally, given that the contracting agents make contracts in light of mutual benefit, this might also mean that SSPDs (sufficiently severely physically disabled) are not adequately

---

<sup>75</sup> Ibid., p. 292.

accounted for with respect to moral matters. (Recall that SSPDs have less to offer than what they need to receive in return.)

But I think that these problems are easily accounted for by the fact of contracting agents being placed behind a veil of ignorance. We can just stipulate this: Any being with moral standing is to be represented somehow by the contracting agents behind the veil of ignorance. In particular, I think that we can use some kind of trustee account. I will characterize a trustee (partially, but crucially) in this way: In virtue of reasoning as a trustee, one considers only the interests of who they are representing. They are not considering their own interest, nor the interest of others, but only the interests of who they are representing.<sup>76</sup>

This is how I think that this trustee account be further characterized and applied. In virtue of having moral standing, you are represented via a trustee. Now, although the contracting agents are, in fact, human beings, they do not know whether they are human beings or not upon the veil being lifted. Suppose that there are 6 billion human beings and something like 100 billion non-human animals with moral standing. Then certainly the contracting agents do not know whether they are a human being or whether they are a de facto trustee for one of the 100 billion non-human animals with moral standing. As such, they reason that it is 100/106 probable that they are a trustee of a non-human with moral standing. So, this should make apparent how non-human animals, SLRC's, and non-human animals are taken into consideration; they are taken into consideration via a trustee. As for SSPD's, they are able to reason just as anyone else. As such, they need no trustee. Furthermore, recall that by being placed behind a veil, a contracting agent, although she knows she has cognitive capacities, does not know whether she is an SSPD or

---

<sup>76</sup> Note that this characterization may depart drastically from how some, or most, people conceive of lawyers. They might think of lawyers as representing their own clients with always their own self-interest in mind, or more minimally, this: Whenever there is a conflict in interest between themselves and their clients, they always choose themselves as a priority. In the case of trustees, just so long as one is a trustee, one is only considering the interests of their client; if there is a conflict, the trustee does not even consider it.

not. It should be apparent now how the outliers are represented by the contracting process according to my veil-based account.

#### *4.4.2 A solution to the gangster-inclusion problem*

A veil of ignorance account directly solves the gangster-inclusion problem. From behind a veil, one does not know whether one is a gangster or not. One does not know her strengths and weaknesses. As such, one will not make agreements via leveling threats against others, as one does not know whether such threats are even remotely legitimately credible. Without such threats, the presence of gangsters among the contracting agents does not affect the terms of the contract.

#### *4.4.3 A solution to the numbers problem*

Recall again from chapter 2 the numbers problem. An ethical theory must be able to show that it is morally permissible to save five people from each suffering harm, of intensity  $x$ , thereby allowing one person from suffering harm, of intensity  $x$ .

According to my veil of ignorance account, the contracting agents would be situated such that they don't know whether they are one of the seven involved in a numbers-problem style case. First, note that I say "seven" because not only would the contracting agents consider the five and the one possibly suffering harm, but also the agent who must make the choice.

Now, what would the contracting agents agree to in numbers-problem style cases, and why? A contracting agent would think: Well, I might be the one person in the group of one. In that case, I would want it to be the case that the acting agent saves me over the five. But the contracting agent would also think: Well, I might also be a person in the group of five. In that case, I would want it to be the case that the acting agent saves the five over the one. A contracting agent also considers how much harm is experienced by each person. If the harm experienced by the one is great enough and the harm experienced by each of the five is small enough, then it seems plausible to think that the contracting agents would agree that the acting agent is permitted to save the one over the five, and perhaps might make this required. But when

harms experienced by each is of equal intensity, then the contracting agents would decide differently. Whether the harm is great or very small, the contracting agents would consider probabilities. And these probabilities may be considered differently. For example, if the relevant set is the five and the one, then the probability of being in the group of five is  $5/6$ . If the relevant set is the five, the one, and the acting agent, then the probability of being in the group of five is  $5/7$ . Because of the probabilistic considerations, and because the harms are equal, and because all other things are equal, the contracting agents, from behind a veil of ignorance, would agree that saving the five is permissible, and perhaps even required. Note that a solution to the numbers problem only requires the conclusion that it is permissible to save the five over the one (the greater over the fewer).

#### 4.5 REPLYING TO PROBLEMS WITH MY ACCOUNT

In this final section, I further motivate my account in numerous ways. In 4.5.1, I explore a certain problem that seems to be unique to an account like mine which has the contracting agents reason according to the principle of maximizing expected utility. In 4.5.2, I motivate my account – both its veil component and its maximizing expected utility component – in terms of its ability to get the right results concerning two kinds of moral cases. In 4.5.3, I explain how my account does not rely on the appeal to analogous moral cases in order to arrive at decisions concerning moral permissibility. Finally, in 4.5.4, I talk about how my account handles certain cases dealing with future generations. In 4.5.5, I make some brief concluding remarks concerning the entire chapter.

##### *4.5.1 One kind of case concerning the contingency and variety of individual preferences*

In this section, I note an already potential problem with my view. Whether or not members of a society (or even a generation of people) experience resentment about certain things, or whether they have a desire for fairness to obtain, is a contingent matter. Furthermore, the proportion of those of a population who have a desire for fairness to obtain may vary, as well. In order to expand on this matter, let's consider the following case. Suppose that we are able to screen any given person on the street on what blood type they have, as well as the relative health of their

organs. Suppose that there are five persons dying from organ failure, each with different organs failing. Suppose that there is one person on the streets that has been targeted as a match. And suppose that there is a kind of superman-doctor that has made the identification, and who upon making the identification, can shoot his dart gun at the match (the person who is a match) and then painlessly remove their organs, thereby killing them painlessly in order to then transplant into the five dying people (and thus saving their lives). What would the contracting agents decide, where each would assign the probability  $1/6$  of being in any of the positions under consideration? (For now, we will not consider the possibility of being the superman-doctor.) Well, what they would agree to will be contingent on other factors, such as the proportion of a sense of fairness throughout society.

Suppose that all members of society had a sense of fairness such that they would attach a great amount of negative utility to a superman-doctor acting in this way. Actions like these would be significant frustrations of each individual's desire for fairness. (Here, I'm using "desire for fairness" in the more specific way mentioned above – namely, "desire that persons' rights are not violated (according to a conception of rights).") More specifically, the utility calculation would not be merely something like the sum of 5 units of positive utility (five people being saved from death) and 1 unit of negative utility. Also, there would be a considerable amount of negative utility upon people knowing that these actions occurred.<sup>77</sup> Additionally, the recipients of the organs would experience some sort of guilt perhaps, and minimally a sense of unfairness at play. At the same time, cases like Trolley would be ones where I think that there would be no sense of guilt or feelings of unfairness. The utility when summed in the Trolley case would be such that

---

<sup>77</sup> To support this point, just really think about a society in which it was permissible for such actions to occur. Yes, in the society in which we live, we have to cope with the fact that premeditated acts of violence occur. But these acts, on the large part, are not tolerated; and part of the basis of their not being tolerated is that they are not permissible, to the point that we have laws against them. Now, imagine a society in which such acts are, in fact, permissible. Furthermore, the law allows them. As such, many people would feel uncomfortable (I'll speculate) at the fact of the such acts being permissible and lawful.

PME would require that one pull the lever (thereby killing one on the alternative track in order to save the five on the original track).

Now let's consider a society in which no one has a sense of fairness, or more poignantly, no one has a desire that their conception of fairness obtains. These agents do not experience guilt in the same way either. In this society it appears that the only considerations to take into account in the superman-doctor case in assigning utilities are the benefits of life to the five recipients of the organs, and the cost to the one who is killed, where the cost is the loss of her life. Perhaps the superman-doctor finds her work grizzly. And any witnesses to the dart gun incident might find this grizzly as well. Perhaps these costs should be taken into account, but when summed up against the benefits gotten by the recipients, it will not be enough to deem the superman-doctor's actions impermissible. But this I think might be the intuitive result. Or, more minimally, I don't think that this is *that* counterintuitive. If all members of society are such that their desire for fairness to obtain is so apparently absent, then why think that the one who has been used for organ transplantation has been wronged? Besides, this person, just as everyone else, does not have a desire for fairness to obtain.

This last point concerning conceptions of rightness, or moral desert, and of what's unfair should be expanded on. In a society in which every member shares the same conception of fairness (and of rightness and desert), and where an action does not frustrate any one's desire that their conception of fairness obtains, it is hard to explain how any wrong-doing (or: unfairness) is at play. Of course, in our society, many of us have the desire that the one healthy person is not treated by the superman-doctor in such a way. But according to my account, this explains why it is wrong in our society for the superman-doctor to act in that way.

Things become trickier if we consider a society as follows: Only 1% of the population has a desire for fairness to obtain. And furthermore, the one who is the candidate organ donor is part of this 1% group. Reasoning from behind the veil of ignorance, the contracting agents know that it is highly unlikely that they are this person. And just with any variation on this kind of 5

versus 1 kind of case, they will know that it is more probable that they need to receive a transplant than be someone used for a transplant. If the superman-doctor carries through with killing the one donor, then this will only frustrate the desire for fairness of a few (the 1% group). These kinds of psychological costs will almost surely not outweigh the benefits gotten by the five recipients. Additionally, we have to consider that in a society there are going to be considerably more people dying from organ failure. Thus, there is a frequency of these 5 versus 1 cases; there are many dying from organ failure, and there are a considerable number of people who can be donors. So, this amplifies the benefits that can be gotten by deeming such superman-doctor practices permissible.

And yet one might say that this is a counterintuitive result – that in such 1% societies, as I will call them, it is permissible for the superman-doctor to proceed with the killing and the transplant, especially when the donor is someone who does have a desire for fairness to obtain. I agree that it is counterintuitive. I will now explain how this counterintuitive result would not be the case – that in such 1% societies, the contracting agents would deem the superman-doctor's action, in fact, impermissible. (Note that my explanation will actually show such a practice to be impermissible in a society where no one has a desire for fairness to obtain – i.e., the society I discussed previous to the 1% society.) Even if 99% percent of the population does not have a desire for fairness to obtain, it still is the case that their psychology is such that they would fear that they be the victim of a superman-doctor's actions. More specifically, they might care very little whether this happens to other people, but it is reasonable to think that they would fear this happening to themselves. Although individuals in this society would know that it is more likely that they be a recipient of organ transplant than being the donor themselves, they might still very much fear dying in such a way – that is, by being randomly (that is, from their perspective) shot by a dart and then killed by a superman-doctor. There is great disutility in a society filled with individuals having this fear. As such, the contracting agents would reason that the expected utility of deeming such a practice (under circumstances of this very specific case) to be

impermissible. (Note that the psychological costs of fear that I'm discussing are not only "weighty", but that they are so weighty that they defeat countering costs such that the super-man doctor practice is impermissible.)<sup>78</sup>

Let me now quickly summarize this section. I've considered certain possible problems with the contracting agents reasoning according to the principle of maximizing expected utility. Because the contracting agents do not take into account a crude account of utility, but instead a very rich one based on the psychologies of people, the results gotten by the contracting agents reasoning according to PME align with our shared intuitions.

#### *4.5.2 Intuitions concerning self-gain at the expense of others (e.g. fraud), and aiding others at the expense of oneself (e.g., famine relief)*

Morality often involves how one ought to act in the prospect of gaining at the expense of others. And it also involves how one ought to act in light of benefiting others even when doing so decreases the level of one's well-being. There are various cases in which we have shared intuitions that self-gain at the expense of another is morally impermissible. And there are other cases in which we have shared intuitions that benefitting others at our expense is morally required. Thus far, I have only considered cases where the moral agent must decide between harming one in order to benefit others (e.g., Trolley Case and Transplant Case). Although I have defended my account with respect to these cases, showing that my account delivers results that cohere with our shared intuitions, I have not said anything of the cases mentioned above – where the agent has to decide whether to harm another for personal gain, or whether to benefit another at the expense of loss to self. In this subsection, I explain how the contractarian view that I endorse also endorses actions cohering with our shared intuitions with respect to these cases.

Recall earlier in the chapter my motivation for the veil of ignorance using the example of slavery. The practice of slavery is one in which, generally, some people gain at the expense of

---

<sup>78</sup> There are, though, hypothetical cases where the superman-doctor's actions would be permissible – for example, in a society where no one has such fears, or where they fear being a member of the five more than being the one sacrificed.

others. And, in general, we find these activities to be morally wrong. I will suggest that according to a veil-based theory, another activity of self-gain at the expense of others is deemed to be, in general, wrong. This activity is fraud. The reason that I think that explaining fraud to be wrong is that fraud sometimes involves some kind of consent. In acting fraudulently to another, it is still possible for one to follow through with the agreement made with another. (Here, think of pyramid schemes.) This is problematic, as some take consent to be the cornerstone of ethical behavior. Yet if consent is the cornerstone of ethical behavior, then certain kinds of fraud will be deemed to be morally permissible, despite our shared intuitions that such behavior is, in fact, impermissible.

I think that a veil of ignorance account can deliver results which align with our shared intuitions concerning such fraud cases (i.e., cases in which fraud occurs although the terms of what is consented to are met). Now, let's consider the contracting agents placed behind a veil of ignorance. And let's suppose that they are considering the consequences for various individuals upon the permissibility of fraudulent activity (where the terms of agreement are nevertheless satisfied). From behind a veil of ignorance, a contracting agent does not know whether she is the one who gains from such activity or whether she is someone who is the victim of such activity. Although fraudulent activity can be quite lucrative, there are very heavy costs to being the victim of fraudulent activity. As such, the contracting agents would agree that fraudulent activity (as described under these conditions) is impermissible. And so, once again veil-based theories win credence in terms of their power to illustrate the wrongness of an intuitively wrong action.

We can go even further in illustrating the power of a veil of ignorance account (like Rawls'). As mentioned earlier, not only are there shared intuitions concerning the wrongness of acting in certain ways, but there are also shared intuitions concerning the wrongness of *failing* to act in certain ways. For example, many of us think that it would be wrong to fail to save a child from drowning in a shallow pond, when the consequences of saving him would be merely wet

slacks and ruined shoes. From behind a veil of ignorance, it would be impermissible for someone, under these conditions, to fail to act in the above-prescribed way.

Famine relief is a similar, but much more controversial and complex, issue. Let's take the following case. Suppose that we are considering the case of a U.S. citizen who makes \$176,117. Furthermore, their assets exceed their liabilities. Also, they have nothing pressing to pay for in the immediate future. Now, suppose that there are 10 million starving and malnourished children in some region of Africa. It is important to also note that this U.S. citizen has done nothing to cause for these 10 million children to exist in their condition; their plight is in no way to due to her. Finally, suppose that there is very good evidence available to this citizen that (i) there is an agency who if are given \$2,000, they will be able to distribute this money in such a way that 10 children (of the 10 million) will be able to live at least another year without suffering, (ii) the agency's distribution from this contribution will not violate any other person's rights, and (iii) the \$2000 will not save more lives in some other way. Also, suppose that without her help, the 10 children will not be aided by anyone else, and will then die within three months. Intuitively, then, the morally required thing to do for this U.S. citizen is that she donate \$2000 to this agency.

But why? According to a veil of ignorance account, we are to think as follows: Suppose that the citizen does not contribute the money. Then what are the effects? According to supposition, the 10 children will die (and painfully). From behind a veil of ignorance, a contracting agent does not know whether she is the U.S. citizen or whether she is one of the 10 children who are close to death. Although giving up \$2000 is a lot of money for our citizen, it is a much greater loss to lose one's life (especially with such agony and possible years of life ahead). As such, the contracting agents would agree that it is impermissible for the citizen to not donate her money. The importance of this example is this: It is one thing for an ethical theory to be able to get results aligning with our shared intuitions concerning the harming of others, but it is another for a theory to get results aligning with shared intuitions concerning the aiding of others.

There are a number of theories which do not easily get the result of requirement to aid in the case above, and yet a veil-based theory does. The fact that veil-based theories achieve this counts in their favor; being able to arrive at results aligning with our intuitions that at times it is required that we help others at costs to ourselves counts in favor of an ethical theory.

#### *4.5.3 Analogous reasoning*

This subsection does not address an objection to my account. Instead, it directly argues for the power of my account. The argument starts from the following observation: As far as I can tell, Rawls does not discuss the power of the veil of ignorance in not having to rely on analogous reasoning in coming to arrive at moral decisions. Rawls' omission of such a discussion is, I believe, to be unfortunate on his part in advancing his account. In the following subsection, I will explain why a veil-based account need not rely, conceptually, on analogous reasoning and why not relying on such reasoning is advantageous. In making my case concrete, I will appeal to how J.J. Thomson uses analogous reasoning in arriving at normative conclusions.

To start, let's consider Thomson's violinist example. I will take it that the reader is familiar with this example.<sup>79</sup> In Thomson's example, we are to conclude that it is morally permissible for a woman to abort where her pregnancy is due to rape. Furthermore, the permissibility of such an abortion is secured even under the assumption of the fetus being a full (moral) person. Thomson points out that if one were kidnapped to save the dying violinist (in order to be subjected to nine months of bed-ridden live-saving help for the violinist), it would be permissible for one to refuse being subjected to such treatment for the violinist; in other words, it would be permissible for one to refuse to be treated in this way, thereby allowing the violinist to die. The analogy that Thomson makes is this: Just as it is permissible for one to deny help to the ailing violinist, it is permissible for the pregnant woman to deny treatment to the fetus (when the pregnancy is due to rape).

---

<sup>79</sup> Judith Jarvis Thomson. "A Defense of Abortion" in *The Rights and Wrongs of Abortion*. (eds.) Marshall Cohen, Thomas Nagel, Thomas Scanlon.

At this point, I want to discuss the legitimacy of such analogous reasoning. At least in this case, the analogy is contentious. One way to contend it is by showing that there is a relevant difference between the ailing violinist and the fetus. For example, one might say that just so long as the fetus is a person, then there is a relevant difference between the fetus as a person and the violinist, as the violinist has lived a much more fulfilling and complete life than the fetus as a person. Whereas it may not be wrong to deny help to the violinist who has lived a somewhat full and complete life, it may be wrong to deny help to the fetus whose life is far from complete. The veil of ignorance provides a way of bypassing the contentiousness. According to a veil account, the contracting agents are able to make decisions without the appeal to analogous cases. Instead, they are able to examine the case before them without appeal to analogy.

To illustrate this point about direct reasoning behind the veil, consider the example of contentiousness made above. Suppose that there is, in fact, a relevant difference between the fetus and the violinist. Thomson's analogy then is ineffective in determining the issue of whether it is permissible to abort the fetus in the case of rape. An appeal would have to be made a case in which the person attached has lived a similarly unfulfilled and incomplete life as the fetus? But who would that analogous person be? Under what we've stipulated, the person would have to be another fetus. And yet in the case of a similar fetus, our intuitions would not be guided in the way that our intuitions are guided by the violinist case. According to a veil-based account, the contracting agents would examine directly the case of the fetus due to rape. A contracting agent would not know whether she is the pregnant woman or whether she was the fetus. If it is relevant, she also does not know whether she is someone other than the pregnant woman or the fetus. I submit that a decision would be made concerning the permissibility of aborting the fetus in this case, but I will not make a gesture to what the decision would be. What's significant here is this: Suppose that arriving at moral decisions via analogous moral reasoning is contentious and/or problematic, then a veil-based account provides an alternative in which no appeal to analogous is needed in order to arrive at (and determine) moral decisions.

#### 4.5.4 *Future generations*

In this final subsection, I discuss how my account can handle a certain problem which any ethical theory faces. This is a problem that we've partially already discussed, namely the problem of future generations. But here, we have to be careful. By "future generations" we can mean two different things. We could mean "any possible future generation" or we could mean "any actual future generation." This is important in the following sense. Suppose that there is a decision to be made concerning which of two possible future generations should be put into existence as actual. The kind of future generations we would be considering then would be possible future generations. The second sense of "future generations" as "actual future generations" precludes the issue of which generation to bring into existence. Instead, it only considers how such actual future generations are to be treated. For the purpose of this subsection, I only have the latter sense of future generations in mind.<sup>80</sup>

Earlier, I discussed Rawls' account in relation to future generations (as understood in the "actual future generations" sense of the concept). His account was problematic in terms of being able to account for such actual future generations. Because only contemporaries are included within the contracting process, and because contemporaries will know the features of the society that they live in – including the kinds of resources that they have as well as the kinds of moral scenarios that arise in such societies – they will also know that any other societal features presented to them, including moral scenarios of those societies, are ones which they will not have to be members of. But this is problematic, as such contracting agents will discard their importance. As mentioned earlier, the best way to remedy this is to include actual future generations to be participants of the contracting process either by (i) including them directly in the process or (ii) making the contracting agents ignorant of whether they are representatives of such agents or not.

---

<sup>80</sup> For important discussions concerning future generations in the "possible future generations" sense, consult Derek Parfit's *Reasons and Persons*. (1984).

Now, what's important is that those either included or represented are persons who are actual persons, and not possible persons, in this sense: These persons *will* exist. It's not that they may exist, but instead it's that given how things will turn out through the future, they will exist. And given this fact, it is certainly important to consider how they are treated. Let me be more concrete. Suppose the existence of two generations, G1 and G2, where G1 is prior to G2. According to my account, members of both generations are accounted for in the contracting process. This is because each member, whether they exist now or not, are actual persons, in that they either do exist or will exist.

Let me reemphasize how my account deals with actual future generations. Since G2 are actual people in the sense that they will exist, it only makes sense to include them in cases involving the treatment of future generations. Notice that this does not require that the contracting agents consider whether they are required to bring more people into existence, since it is already determined who will be brought into existence (i.e., actual persons who will exist). So, suppose that the contracting agents are determining whether to permit certain practices that will cause in the future the practices of slavery on a wide-scale. According to an account like mine, where a veil and principle of maximizing expected utility is employed, these practices would not be permitted. Because members of G2 are represented in the contracting process, and because one does not know whether one is a member of G2 (or representing a member of G2), one would understand that the expected utility of agreeing to these practices (causing wide-scale slavery) is less than the expected utility of agreeing to not permit such practices.

#### *4.5.5 Conclusion*

I will now give some brief remarks concerning the conclusions of this chapter. In 4.2, I illustrated some main objections to Rawls' view, and in particular, his view with the further assumption of the veil of ignorance. In 4.3, I then responded to these objections by noting any kind of departures that should be made from Rawls' view without the sacrificing of a veil of ignorance. The replies required that contrary to Rawls: (i) the contracting agents are able to

assign utilities according to the preferences and capacities of all the persons affected by a moral case, but that the contracting agents do not know what their particular capacities and preferences are; (ii) the contracting agents make decisions according to the principle of maximizing expected utility, and not maximin; (iii) the contracting agents do not aim to advance primary goods, but instead, once again, aim to maximize expected utility; (iv) the contracting agents, although they are actual people (in that they either exist or will exist), do not know what generation they belong to. Additionally, I have not assumed that the contracting agents are not mutually unconcerned. This, in particular, was important not to assume in order for utility assignments to reflect the rich psychology of actual people. Namely, we often have desires that others are treated and not treated in certain ways. When those desires are frustrated, this amounts to negative utility being assigned. And from this, we can explain why in many societies such actions as the superman-doctor killing an innocent person for their organs is morally impermissible.

In 4.4, I explained how a veil of ignorance account can handle the three problems discussed in Ch.2, which were genuinely problematic for other ethical contractarian theories. And in 4.5, I further motivated my account, showing the power of its combination of employing a veil of ignorance with the rational choice principle of maximizing expected utility.

## CHAPTER 5 – CONCLUSION

In this final chapter, I first make some remarks concerning the conclusions reached in this dissertation. I then discuss some further areas of research concerning the view that I defended in this dissertation. In particular, I take there to be four areas where further research would be productive in light of Rawlsian ethical act contractarianism.

### 5.1 SUMMARY

This dissertation has been an extended argument for a particular contractarian ethical theory, one which I have called “Rawlsian ethical act contractarianism.” Recall that such an account was not offered as a theory of justice, either in the form of justifying the existence of government or in the form of justifying particular principles of distributive justice. Instead, the kind of contractarianism advanced is a theory of the nature of permissible and impermissible actions and practices, and as such, is advanced as an ethical theory.

Recall furthermore that the theory is characterized as follows:

Rawlsian ethical act contractarianism: An action A is morally permissible if and only if, from behind a veil of ignorance, A would be agreed to by all contracting agents, each of which are the way they actually are in real life, except that they are perfect expected individual utility maximizers.

The two major components of Rawlsian ethical act contractarianism is (i) its Rawlsian component of employing a veil of ignorance and (ii) its act component, where the objects of agreement are not rules but instead specific actions.

In terms of the Rawlsian veil of ignorance, I contrasted it with features essential to two contemporary proponents of competing classical contractarian accounts – in particular, the Hobbesian contractarianism developed by Gauthier and a version of Kantian contractarianism developed by Scanlon. I showed that each of these accounts suffer from various problems which

are not endemic to an account employing a veil of ignorance. Nevertheless, a number of theorists have thought the veil of ignorance to be problematic for other reasons. As such, I defended my account against these purported objections.

The second component, the act-based component, was argued for in Chapter 3 by contrasting it with rule contractarianism, and then showing that the latter kind of theory suffered from various problems. In particular, we saw that if rule contractarianism is practically equivalent to act contractarianism, then it will have to endorse rules which violate certain practicality constraints on what counts as practical moral rules. But then in avoiding practical equivalence, rule contractarianism runs amuck by either being indeterminate or by rendering results in specific cases which are implausible. Or, if rule contractarianism is to avoid these charges, then it can always bite the bullet and maintain such rules, but in doing so, it amounts to a kind of rule-worship – i.e., maintain rules despite their application allowing for implausible results.

A note should be made about mutual positive support for act contractarianism and the employment of a veil of ignorance. As I had discussed in Chapter 3, I pointed out that act contractarianism without a veil becomes less plausible. We can imagine certain conflict cases in which agreement about specific actions would not be reached if the contracting agents actually knew their respective positions with respect to all the available alternatives. As such, it may make more sense to agree to rules, since such agreement may occur without examining specific cases involving conflict between agents. But with a veil of ignorance, agreement on specific actions (where agreement is made in light of an action being the one that maximizes expected utility) is possible and will, in fact, occur. Just so long as a contractarian account takes on an act-component, then it should also employ a veil of ignorance. But as pointed out, the employment of a veil has been defended in this dissertation. Furthermore, I have argued that a rule-based account (whether it employs a veil or not) runs into problems. As such, this motivates the idea of

an act-based view. I will now move to the next section, where I suggest further areas of research pertaining to my dissertation.

## 5.2 AREAS OF FURTHER RESEARCH

### 5.2.1 *Avoiding act utilitarianism*

Given that the contracting agents of my account, while being placed behind a veil, reasons according to the principle of maximizing expected individual utility, one might wonder how closely related my view is to utilitarianism. Furthermore, since the objects of agreement are actions, and not rules, one might wonder about the view's relationship to act utilitarianism. Now, obviously the two views are distinct. My view says that actions are permissible just in case they are agreed upon by contracting agents (with certain characteristics in a certain choice situation), whereas act utilitarianism says that actions are permissible just in case they maximize overall utility. But one might wonder why I don't just bypass the contracting agents and instead say that actions are permissible in light of the standard of maximizing overall utility. I will leave this to be further explored, but make some brief remarks nevertheless.

First of all, my dissertation could be thought of arguing for a specific kind of view under the assumption that some kind of contractarianism is the correct form for an ethical theory. Thus, this is one reason not to endorse a direct act utilitarianism. Secondly, there are good reasons to start with contractarianism. The idea that morality is a product of agreement among rational people is attractive. Furthermore, the idea that they arrive at their agreement while behind a veil of ignorance has been defended in this chapter, often via an appeal to arriving at results which align with our shared intuitions. And given that the contracting agents are placed in a choice situation of uncertainty via a veil, it then makes sense to characterize their reasoning/rationality according to the principle of maximizing expected utility. In other words, the reason for endorsing a utilitarian rational choice principle has to do with prior commitments to a contractarian view employing a veil. As such, what's more fundamental to my account is the contractarian and veil component, and not any utilitarian rational choice principle.

### 5.2.2 *What room is there for moral prerogatives?*

I have defended the view that actions are permissible just in case they are agreed upon by the contracting agents while being placed behind a veil. And contracting agents decide upon actions based on being presented specific moral cases. So, when contracting agents are presented a moral case, they agree upon an action from a set of possibilities (or alternatives). What could be worrisome, though, is this: What if the contracting agents are presented with some case in which they agree to some action and yet we think that there is a moral prerogative that makes it permissible for real-life agents to do otherwise?<sup>81</sup>

Take for example the following case: Suppose that upon Alicia asking Bruce out on a date, she must make a choice between taking him to dinner for Tai or for Italian cuisine. For Alicia, both options are equally desirable. But she does not know which he prefers, although she knows that Bruce likes both. It turns out she that takes him to get Tai, when, in fact, he prefers Italian. Has Alicia done anything morally wrong? Of course not; we think that there is a moral prerogative here such that her choosing to take Bruce to either restaurant is morally permissible. But if the contracting agents were presented this case, then they would agree upon the action of Alicia taking Bruce to get Italian. And thus, her action of taking him for Tai would be morally wrong. Well, this gets the wrong result. Namely, we should think that she does nothing wrong either way.

The idea is this: Further research could be done on how a contractarian theory could account for moral prerogatives. Certainly we do not want a contractarian theory to dictate a number of actions which are morally permissible (or: what some may even call “morally indifferent”) as morally impermissible. In general, we want an ethical theory to avoid being too morally demanding. Of course there is a place for an ethical theory to demand aid to others on a moral basis. But this can be taken too far. The idea here is that contractarianism, according to the view that I’m advancing, not be taken too far.

---

<sup>81</sup> See Samuel Scheffler’s *The Rejection of Consequentialism*.

### 5.2.3 *Future generations*

Recall that the contracting agents are actual people with certain idealized characteristics. The idea that merely possible people are included into the contracting process is excluded according to my account. We then have a problem concerning choices affecting future generations.

Suppose that we are to decide to choose between two (or more) possible future generations whose members do not overlap; let's call them G3 and G4. (I've named them G3 and G4 because a related issue was discussed in Chapter 4 in which G1 and G2 were already used for naming purposes.) Now, according to a veil-based account, this can be decided according to one of two ways. First way: Follow the account I've argued for, where only actual persons decide the matter. Now, the first problem with this approach is this: If G3, for example, is decided to be brought into existence, then the members of G3 are in fact actual people, as they will exist. Thus, by including all actual people, G3 is to be brought into the contracting process to decide the matter on whether G3 or G4 will exist. But supposing that this is a genuine decision to make, G3 can't consist of actual people (more specifically, people who will exist). We have a reductio of some sort. Therefore, we should reject the idea that in these matters actual people (which includes future people) will decide the matter.

The other way involves not only actual people, but all possible future people. So, in addition to actual people existing in the present generation, we would have all members from both G3 and G4 included. We could do this by making the present actual generation ignorant of whether they are of the actual generation, or of G3, or of G4. But now we have a problem related to one that Derek Parfit has raised.<sup>82</sup> Suppose that G3 consists of 5 billion members, each living a good life. And suppose that G4 consists of 40 billion members, each living a much worse life, yet one which is still barely worth living. Now Parfit points out that a reasonable consequentialist principle will lead us to choose G4, just so long as the total sum of what makes life worth living

---

<sup>82</sup> See Derek Parfit's *Reasons and Persons*, pp. 351-390.

is greater in G4 than in G3; he calls this the Repugnant Conclusion. For the purposes of contractarianism, we can state the problem in a slightly different way. If we work with the assumption that having a life which is barely worth living is better than not having lived at all, then we could imagine the contracting agents deciding (as expected utility maximizers) to bring G4 into existence. Now, this point can be generalized as follows: We could imagine the contracting agents, if possible persons are included, to bring about massive procreation policies. To use Parfit's terminology, I find this to be repugnant.<sup>83</sup>

### 5.3 CONCLUDING REMARKS

Let me now just make some brief concluding remarks. This dissertation has defended Rawlsian ethical act contractarianism as an ethical theory (more specifically, a theory of permissible action). The two major components of it are the Rawlsian veil of ignorance and the act, as opposed to rule, component. After defending these components, I've in this chapter suggested further areas of research, namely areas dealing with the distinctness of this contractarian view from utilitarianism, with the availability of endorsing moral prerogatives, and also problems concerning future generations.

---

<sup>83</sup> Here's my tentative solution: Of the two major ways to decide bringing future generations into existence (where we are capable of determining these kinds of things in an organized way), we should avoid the inclusion of possible people in the contracting process. We should include only actual people, but furthermore make the following restriction: In cases where we're choosing between G3 and G4, and where one of the two will actually exist, we restrict all members from both from choosing. Suppose that currently G1 exists and that it is already determined that G2 will exist. Then both G1 and G2 are to decide the matter. They are aware that they are not members of G3 or G4. Because of this we will have to tweak the motivation conditions of the contracting agents. They are to imagine what it would be like being a member of G3 in contrast to being a member of G4. I will speculate that since they will find the terms for existing under G3 are better, they will decide that G3 is to be selected over G4. It is obvious that this is far from conclusive, and so I welcome any kind of further research on this issue – namely, the issue of how a contractarian veil-based account is to deal with the moral matter of choosing between possible future generations. Obviously, this issue is not decided by my tentative solution. Instead, further research on this should be done. It would be rather exciting to see the various solutions that are offered in light of the problems that occur about future generations when a veil of ignorance is employed.

## BIBLIOGRAPHY

- Becker, Lawrence C. "Reciprocity, Justice, and Disability." *Ethics*. 116 (2005): 9-39.
- Brandt, Richard B. *A Theory of the Good and the Right*. Oxford University Press; Oxford: 1979.
- Cudd, Ann. "Contractarianism." *Stanford Encyclopedia of Philosophy*. 2000.
- Dancy, Jonathan. *Ethics Without Principles*. Oxford University Press; New York: 2004.
- D'Agostino, Fred. "Contemporary Approaches to the Social Contract." *Stanford Encyclopedia of Philosophy*. 1996.
- Daniels, Norman. ed. *Reading Rawls*. Basic Books; New York: 1975.
- Dworkin, Ronald. "The Original Position." In *Reading Rawls*. Norman Daniels (ed.). Basic Books: 1975.
- Gauthier, David. *Moral Dealing: Contract, Ethics, and Reason*. Cornell University Press; Ithaca, New York: 1990.
- \_\_\_\_\_. *Morals by Agreement*. Clarendon Press; Oxford: 1986.
- Hampton, Jean. *Hobbes and the Social Contract Tradition*. Cambridge University Press; Cambridge: 1986.
- Hare, R.M. *The Language of Morals*. Oxford University Press; Oxford: 1952.
- \_\_\_\_\_. "Rawls' Theory of Justice." In *Reading Rawls*. Norman Daniels (ed.). Basic Books: 1975.
- Harsanyi, John C. *Essays on Ethics, Social Behavior, and Scientific Explanation*. D. Reidel; Dordrecht, Holland: 1976.
- Hart, H.L.A. "Rawls on Liberty and its Priority." In *Reading Rawls*. Norman Daniels (ed.). Basic Books: 1975.

- Hooker, Brad. *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*. Oxford University Press; Oxford: 2000.
- Kagan, Shelly. *Normative Ethics*. Westview Press; Boulder, Colorado: 1998.
- Kamm, Frances Myrna. *Creation and Abortion*. Oxford University Press; New York: 1992.
- Kavka, Gregory. *Hobbesian Moral and Political Philosophy*. Princeton University Press; Princeton, NJ: 1986.
- Morris, Christopher. "Moral Standing and Rational-Choice Contractarianism." In ed. Peter Vallentyne. *Contractarianism and Rational Choice*. Cambridge University Press: 1991.
- Nagel, Thomas. "Rawls on Justice." In *Reading Rawls*. Norman Daniels (ed.). Basic Books: 1975.
- Narveson, Jan. *The Libertarian Idea*. Temple University Press; Philadelphia: 1988.
- Nozick, Robert. *Anarchy, State, and Utopia*. Basic Books; New York: 1974.
- Parfit, Derek. *Reasons and Persons*. Oxford University Press; Oxford: 1984.
- \_\_\_\_\_. "Innumerate Ethics." *Philosophy and Public Affairs*. 1978.
- Rawls, John. *Lectures on the History of Moral Philosophy*. Barbara Herman. ed. Harvard University Press; Cambridge, MA: 2000.
- Rawls, John. *A Theory of Justice*. Harvard University Press; Cambridge, MA: 1971.
- Reibetanz, Sophia. "Contractualism and Aggregation" *Ethics*. 1998
- Scanlon, T.M. *What We Owe Each Other*. Harvard University Press; Cambridge, MA: 1998.
- Scheffler, Samuel. *The Rejection of Consequentialism*. Clarendon Press; Oxford: 1982.
- Silvers, Anita, and Francis, Leslie Pickering. "Justice Through Trust: Disability and the 'Outlier Problem' in Social Contract Theory." *Ethics*. 116 (2005): 40-76.
- Smart, J.J.C., and Williams, Bernard. *Utilitarianism: For and Against*. Cambridge University Press; Cambridge: 1973.
- Taurek, John. "Should the Numbers Count?" *Philosophy and Public Affairs*. 1977.

Thomson, Judith Jarvis. *The Realm of Rights*. Harvard University Press; Cambridge, MA: 1990.

\_\_\_\_\_. "A Defense of Abortion." In *The Rights and Wrongs of Abortion*. Marshall Cohen, Thomas Nagel, Thomas Scanlon. eds. Princeton University Press; Princeton, NJ: 1974.

Vallentyne, Peter. ed. *Contractarianism and Rational Choice*. Cambridge University Press; Cambridge: 1991.

\_\_\_\_\_. "Contractarianism and the Assumption of Mutual Unconcern" in *Contractarianism and Rational Choice*. (ed.) Peter Vallentyne. Cambridge University Press; Cambridge: 1991.

## VITA

Joel Dittmer was born, raised, and educated in Missouri. Born in 1980, he was educated continuously in Missouri public schools, from elementary school to the University of Missouri, graduating in 2002 with a B.S. in mathematics and in 2005 with an M.A. in philosophy. He will graduate in 2010 with a Ph.D. in philosophy from the University of Missouri. He is interested in a number of things professionally (ethical theory, applied ethics, the history of philosophy, and philosophical methodology), and enjoys nature walks, rock n' roll, and the discussion of ideas with anyone.