

Contractarianism and ‘Moderate Morality’

by

Vaughn Bryan Baltzly

This thesis is submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

IN

PHILOSOPHY

APPROVED:

William FitzPatrick, Chair
Harlan Miller
James Klagge

June, 2001
Blacksburg, Virginia

ABSTRACT

In his book *The Limits of Morality*, Shelly Kagan claims that contractarian approaches to ethics are incompatible with our common, everyday, “moderate” morality. In this thesis I defend a version of contractarianism that I believe leads to both deontological constraints and options; i.e., to a genuinely *moderate* morality. On my account, the parties to the agreement are conceived of as being motivated not only to promote self-interest, but also to formulate a code of ethics that gives proper respect to their moral status as persons. If such a picture of the bargainers’ motivations is defensible, as I believe it is, then the ‘moderate’ *may* in fact have recourse to contractarianism in her defense of everyday morality, for – as my thesis argues – bargainers that are thus motivated will arrive at a moderate morality.

TABLE OF CONTENTS

Table of Contents	iii
Acknowledgments.....	iv
Introduction	1
Chapter One: The Challenge to Moderate Morality	3
I. <i>Dramatis Personae</i> : The ‘extremist’, the ‘minimalist’, and the ‘moderate’	3
II. Kagan’s two challenges	5
III. Why contractarianism?	7
Chapter Two: A Contractarian Response to the Challenge	11
Chapter Three: Modified Constraints	14
I. The self-interest-maximizing motive and restrictions on the pursuit of individual good.....	15
A. The prisoners’ dilemma.....	15
B. The trade-off.....	17
C. Minimizing the trade-off	18
II. The personhood-respecting-motive and restrictions on the pursuit of individual good.....	20
A. Inviolability	21
B. Morality’s ‘giving proper expression to’ persons’ inviolability.....	24
C. The trade-off.....	26
D. Minimizing the trade-off	27
E. Thresholds.....	29
III. The shape of the modified constraints	29
Chapter Four: Options	31
I. Self-interest and options: the appeal to cost.....	31
II. The personhood-respecting motive and options.....	32
III. Concluding remarks.....	34
Bibliography	36
Vita.....	37

ACKNOWLEDGMENTS

I would like to thank my committee members Harlan Miller, James Klagge, and William Fitzpatrick for all their help with this thesis. I must especially express my gratitude to Drs. Miller and Klagge for the promptness with which they were willing to read my material, and to my advisor Dr. Fitzpatrick for all his patient advice and comments. Thanks also are due to the entire Virginia Tech philosophy faculty and to the graduate students, from whom I could always expect a helpful conversation when needed, a pleasant diversion when desired, and most importantly, someone nearly forcing me to get some work done when such coercion was not desired, but certainly needed. Especially helpful in this regard was Jesse Ehnert, with whom I had many conversations that were somehow simultaneously helpful and pleasantly divergent.

INTRODUCTION

In his 1989 book *The Limits of Morality*, Shelly Kagan claims that our ordinary, everyday morality – which he dubs ‘moderate morality’ – is untenable and must be abandoned in favor of a more stringent ethical code, which he calls ‘extremism.’ This thesis represents an attempt to defend moderate morality against this claim by appealing to a contract-based approach to ethics. In taking on Kagan’s general charge that moderate morality is incoherent, I adopt two of his specific challenges from chapter four as a starting point. These challenges are as follows: first of all, if the moderate wants to establish the existence of constraints, she must also show that these constraints and their principled exceptions (required, for example, if we are to allow most cases of harming others in self-defense) can be generated from a common principle. And secondly, any moderate hoping to appeal to contract approaches to ethics as a way of defending ordinary morality has the further burden of demonstrating that such an approach does not in fact offer support for extremism.

In order to defend a contract-based moderate morality against both Kagan’s more specific and more general challenges, I must show that my account leads to three things: an acceptance of a ‘*pro tanto*’ reason to promote the greater overall good, deontological constraints that admit of principled exceptions, and options *not* always to perform the action that results in the greater overall good. I am able to ground all these things in my contract approach because I construe my bargainers’ motivation as being twofold: on the one hand, they are motivated to maximize self-interest, as are the bargainers in many other prominent contractarian accounts. But on the other hand, they are also motivated to formulate a set of rules that gives proper expression and recognition to their *moral status as persons*. The moral status that the bargainers want their contract to recognize has at least three dimensions: on such a conception, persons are simultaneously *worthy of concern, ends-in-themselves, and inviolable*.

After setting up Kagan’s challenges in some detail and motivating my decision to use contractarianism as a response to these challenges in chapter one, and after offering a brief sketch of my bargainers’ twofold motivation in chapter two, I go on to give a somewhat detailed account of the way such a contractarian scenario might yield a

moderate morality sufficient to meet Kagan's challenges. Chapter three explores a way in which constraints – constraints which are genuinely deontological, yet which nevertheless admit of principled exceptions and thresholds – might arise from the bargainers' self-interest-maximizing motive and their concern that the moral system they adopt give adequate recognition to their inviolability. Chapter four then proceeds to ground options in the bargainers' desire to formulate a set of rules that properly recognizes their status as ends-in-themselves. Attention is also given to the fact that acceptance of the *pro tanto* reason to promote the greater overall good emerges from the bargainers' belief that persons are 'worthy of concern.'

It should be noted at the outset that I do not take this to be anything like a complete presentation of a contractarian account. Rather, it is more a gesture toward what might serve as a fully-articulated contractarian response to Kagan's challenge in *The Limits of Morality*. Nor is the contractarian account towards which I gesture taken to be anything like a full account of morality. Instead, I wish to follow several recent prominent contractarians¹ in pointing out that such an account only deals with one specific – albeit, very important – aspect of morality: that part of ethics which concerns, in T. M. Scanlon's phrase, "what we owe to each other."² Nevertheless, I hope that this study can make a useful contribution to the ongoing discussion regarding both contractarianism and our ordinary, everyday, 'moderate' morality.

¹ Cf. Rawls, pp. 12 – 14, and Scanlon, pp. 6 –7.

² Scanlon, p. 7.

CHAPTER ONE: The Challenge to Moderate Morality

I. *Dramatis Personae*: The ‘extremist’, the ‘minimalist’, and the ‘moderate’

The Limits of Morality is Shelly Kagan’s challenge to what he terms ‘moderate morality.’ The position of the ‘moderate’ is characterized by her acceptance of three things: a *pro tanto* reason³ to promote the greater overall good, *constraints* on what may be done, either for the sake of promoting the greater overall good or for the sake of promoting personal good, and *options* (sometimes called ‘prerogatives’) *not* to perform the action which will result in the greater overall good. The moderate is distinguished from the ‘minimalist’ on the one hand and the ‘extremist’ on the other. The minimalist is one who, though he may believe that there are some constraints on one’s pursuit of his own individual good, does not accept any *pro tanto* reason to promote the overall good, and may not accept any constraints on the pursuit of individual self-interest at all.⁴ The extremist, meanwhile, is one who, while he may or may not believe in constraints on what may be done in order to maximize the greatest overall good, does *not* believe in options to favor personal interests over promoting the greater overall good.

Kagan’s goal in *The Limits of Morality* is to apply pressure to the moderate (whom he understands as giving voice to our ordinary, everyday, common-sense morality), showing her that she must give up her untenable middle-of-the-road position, as it does not withstand scrutiny, and embrace either minimalism or extremism. (Kagan himself favors the extremist’s position, but it is not his task in this book to argue positively in favor of this view.) Kagan’s main strategy in pressuring the moderate is to subject her justifications of both constraints and options to severe criticism. The defense of options undergoes particularly heavy attack, for, as we have seen, the commitment to options is unique to the moderate. For though the minimalist likely does not accept the *pro tanto* reason to promote the greater overall good, the moderate does share this acceptance with the extremist. Similarly, the moderate’s insistence on constraints might

³ According to Kagan, a ‘*pro tanto*’ reason to promote the good is one that “has genuine weight, but nonetheless may be outweighed by other considerations . . . [A] *pro tanto* reason is a genuine reason – with actual weight – but it may not be a *decisive* one in various cases.” (Kagan 1989, p. 17)

⁴ "Note that a variety of sharply diverging positions will fall within the minimalist camp, including egoists (who believe that one is never required to sacrifice overall self-interest), nihilists (who believe that

be held in common with the minimalist or even the extremist. Kagan explicitly discusses the perfectly coherent (even if somewhat surprising) possibility that an extremist morality might feature constraints:

According to the extremist, morality does not straightforwardly require you to choose that act, *whatever* it is, which can be reasonably expected to lead to the best consequences. [Such a view would constitute a consequentialist morality.] Rather, it restricts your choice to those acts not otherwise forbidden. However, the extremist's claim itself is neutral on the issue of whether morality ever *does* forbid an act which would lead to the best consequences . . . [Suppose] constraints *can* be justified. Such a result is still compatible with the extremist's claim, and the extremist will continue to demand far more of an agent than the moderate . . . If there are constraints, then you are required to perform the optimal act among those acts which do not violate them.⁵

Yet the moderate's position is different even from that of the extremist who believes in constraints because the moderate also believes in options. This is just to say that "the extremist accepts, and the moderate rejects, a *general* requirement to promote the good."⁶ This claim that the moderate rejects the general requirement to promote the good is not to be confused with Kagan's claim that the moderate also accepts the *pro tanto* reason to promote the good; it is only that her acceptance of this *pro tanto* reason is tempered by her acceptance of options.

So it is the moderate's defense of options that bears the brunt of Kagan's attacks on moderate morality. Kagan believes it is the moderate's commitment to these options that makes her position so hard to defend, and that if the moderate's justification of options does not hold up in the face of his examination, she will be forced either to retreat into minimalism or to accept extremism.

So much, then, for the general contour of Kagan's overall attack on moderate morality. Let us examine in detail two of the more specific challenges he offers to the moderate.

everything is morally permitted), and extreme libertarians (who recognize the validity of constraints, but deny that there is a moral requirement to provide aid)." (Kagan 1989, p. 5).

⁵ Kagan 1989, p. 8.

⁶ Kagan 1989, p. 9.

II. Kagan's two challenges

As part of this overall, two-pronged critique of moderate morality, Kagan devotes chapter four of *The Limits of Morality* to a sustained examination of the moderate's endeavors to justify constraints on intending harm. Along the way, he points to several difficulties any such endeavor must face. The first one is the troublesome case of intending harm in self-defense. As Kagan points out, most – if not all – moderates who believe in a constraint against intending harm believe that legitimate exceptions to this constraint arise when intending harm becomes necessary for self-defense. But this is problematic for the moderate: how can these exceptions be justified? The constraint obviously must be modified. But as Kagan writes, “[M]aking the modification may be easy; defending it is not . . . for if the exception is not to be *ad hoc*, it will have to be compatible with the grounds for the constraint itself.”⁷ He goes on to offer the moderate a very specific challenge:

[T]he question is whether the moderate can defend the suggestion that the reasons that normally oppose intending harm fail to be generated when I face Schmidt [an aggressor]. Obviously, the moderate cannot simply assert an exception for the guilty: he needs to show how the exception flows naturally from the account offered of how and why those reasons are normally generated in the first place. Thus the success of such a defense will depend on the particular account offered.⁸

I will call this Kagan's *first* challenge: the charge that any fully-articulated 'moderate' morality must, when necessary, allow for principled exceptions to constraints that are founded in the same principles that ground the constraints themselves.

However, Kagan does not want to claim that no such account could be offered. He immediately recognizes that either of two 'indirect' moral theories might provide just such a justification: the “two-level” (roughly, “rule-utilitarian”) approach, and the “contract” approach. Yet, as Kagan quickly points out, he has already demonstrated in chapter one of his book that either of these approaches serves ultimately to support an extremist morality.⁹ So Kagan's first challenge still looms large before the moderate – in fact, it is made even tougher – when it is coupled with this charge. I will call this

⁷ Kagan 1989, p. 134.

⁸ Kagan 1989, p. 135.

⁹ See Kagan 1989, pp. 32 – 46.

Kagan's *second* challenge: namely, the charge that, since "the moderate . . . cannot turn to these theories now in support of self-defense, [as] such an appeal would undermine the moderate's larger defense of ordinary morality,"¹⁰ she must either avoid appealing to two-level or contract approaches, or else show that such an appeal will not – despite Kagan's arguments to the contrary – lead to extremism.

Kagan offers these two specific challenges as part of his larger, overall attack on moderate morality. In this thesis, I shall take up these two specific challenges as part of a larger, overall *defense* of moderate morality. Specifically, I want to explore the resources offered by a particular contract-based approach to morality that I believe not only grounds constraints and their principled exceptions in a common source – thereby meeting Kagan's first challenge – but also leads to options, and thus to a genuinely 'moderate' morality – thereby meeting Kagan's second challenge as well. By offering a specific response to two of Kagan's specific challenges in *The Limits of Morality*, I hope to point towards a more general defense of moderate morality in response to the more general criticism of it that is motivating Kagan's two challenges. That is, I hope to argue for a consistent, contract-based, moderate morality in the face of Kagan's charge that any form of moderate morality is incoherent and collapses into one of its two, more extreme alternatives.

Some may wonder if these two challenges of Kagan's are even worth taking up. Many people, for example, may want to dismiss Kagan's first challenge as unreasonable: it simply is *not* the case, they might say, that one must show that constraints *and* their exceptions flow from a common principle. Thinking otherwise places too much of a burden upon moral philosophy; we should not expect it to be as unified and cohesive as are our theories in, say, physics. However, I think that it is worthwhile to take up this challenge. If we can do so successfully, we will have met Kagan's challenge on his own terms, rather than having simply avoided it. Furthermore, in the process, we will have arrived at a more robust, fully-articulated understanding and defense of our ordinary, moderate morality. While it indeed may *not* be the case that such a theory crumbles if it *cannot* meet Kagan's first challenge head on, it certainly seems that such a defense – if

¹⁰ Kagan 1989, p. 135

successful – will do much to increase its plausibility. I will try to provide just such a defense.

Of course, once we decide to take up Kagan’s first challenge, it becomes imperative that we take his second challenge seriously too: we had better either avoid appealing to a social-contract scenario to ground our constraints, or else be able to take Kagan up on this challenge and show that our contractarian story does *not* lead to extremism, but leads instead to a genuine moderate morality complete with constraints and options. And again, though the moderate’s inability to meet Kagan’s second challenge may not necessarily be fatal to her position, her successful response to this challenge *will* do much to increase the plausibility of moderate morality in the eyes of the extremist.

III. Why Contractarianism?

It might seem that by appealing to contractarian theories as a means of replying to Kagan’s critique of moderate morality, I am setting before myself a needlessly difficult task. For, while it may be true that the contract approach does provide a promising way to ground constraints and their principled exceptions in a common source (as Kagan himself suggests on page 135), it also has the off-setting burden of presenting us with Kagan’s second challenge – I must now additionally show that the contract approach leads to options. So perhaps a word needs to be said about my decision to utilize contractarianism in meeting Kagan’s challenges.

I have already more than once suggested that contractarianism seems, at first glance, to provide the best means of meeting Kagan’s first challenge. But why is this? I believe it is due to an essential feature of all contractarian accounts: the fact that in contractarian scenarios, bargainers will seek to minimize the ‘trade-offs’ that accompany the implementation of the social contract. That is, bargainers seeking to adopt a contract for the sake of lessening the evils accompanying the state of nature do so even as they foresee that certain trade-offs will occur. For while adopting the contract will bring about protection from a great number of the evils the bargainers would encounter in the state of nature, this protection comes at the price of certain new threats that are sure to arise in life under the contract. The bargainers accept the threat of these new evils because they prefer the trade-off; still, though, they will seek to make this trade-off as slight as

possible. And it is precisely this desire to minimize such a trade-off that makes contractarianism such a promising means of justifying the principled exceptions to constraints on harming others required by some cases of self-defense. For it seems obvious that if a set of bargainers agrees to certain constraints on behavior because, for example, adherence to those constraints is expedient in fulfilling their desire to maximize overall well-being, and if the bargainers also can reasonably foresee that on a few particular occasions, their very adherence to those constraints should happen to threaten or undermine overall well-being, then it seems an exception to those constraints would be warranted on those particular occasions. Furthermore, as we have just seen, this exception is grounded in the very principle that grounded the constraints in the first place – namely, the bargainers’ desire to maximize overall well-being. Any constraints agreed to by a set of bargainers, it seems, will thus be qualified in such a way as to allow for certain principled exceptions. This attribute – call it ‘principled defeasibility’ – would therefore seem to be a necessary feature of *any* successful contractarian account; no wonder, then, that contractarianism seems a promising route to take when trying to meet Kagan’s challenges. And certain cases of self-defense certainly provide a good illustration of this: we would not expect our bargainers to remain constrained to adhere to restrictions against harming others, even if such measures might be their only means of protecting themselves from the attacks of people who are breaking these very restrictions. Such an exception would be warranted as a result of the bargainers’ desire to minimize, in this case, the trade-off posed by the threat of *free-riders*. In this respect, then, contractarianism seems to be the best choice among the available alternatives; neither Kantian theories, virtue theories, nor divine-command theories seem to offer as plausible a justification of principled exceptions to constraints as contractarianism can.¹¹ I will have more to say about the precise way in which my contractarian account will ground constraints and their principled exceptions in a common source in chapter three below.

¹¹ Act-consequentialist theories, of course, offer an easy means of justifying exceptions to *secondary* rules, or ‘rules-of-thumb’, but of course these are not *constraints* in the relevant sense of the term. At any rate, consequentialist theories are of necessity extremist theories – namely, extremist theories that do not recognize any constraints on the *pro tanto* reason to promote the greater overall good. Hence the specification that contractarianism is the best candidate, among the *available* alternatives, for the task of providing a moderate morality that will meet Kagan’s two specific challenges and hold up against his general charge that options cannot be defended.

But what of Kagan's second challenge? Might not the difficulty of defending options *as well as* constraints, in the face of Kagan's charge that contractarian accounts will lead always to extremism, be sufficient to outweigh the relative ease with which I might hope to defend constraints and their principled exceptions? Would not this dissuade me from trying to appeal to contractarianism as a reply to Kagan's general challenge to moderate morality? Again, I want to argue that special particular features of contractarianism make it the natural choice for anyone hoping to defend moderate morality, just as we have seen it is the natural choice for anyone hoping to ground constraints and their principled exceptions in a common source. To see this, simply consider the (what I take to be misguided) way in which Kagan argues that the contract approach leads essentially to extremism.

Kagan argues for this conclusion by pointing out that rational bargainers motivated to maximize their self-interest or well-being will never be inclined to accept options not to promote the overall good: such options would likely decrease the average overall good, and thus lower each individual bargainer's chances of maximizing his or her own personal good. This argument, though, neglects the possibility that there may be *two kinds* of goods the bargainers might be concerned with at the bargaining table: what I will term for present purposes *natural goods*, and what I will term *moral goods*. Before proceeding, let us quickly get clear on how I will be using these terms. Natural goods are goods like health, pleasure, well-being, and prosperity. They are the sorts of goods that can be sensibly thought of as being promoted within a moral system. *Moral goods*, on the other hand, are not the sorts of things that might be *promoted* by a moral system; rather, they are the sorts of things that can only be reflected *in*, or recognized *by*, a moral system. We may legitimately ask of natural goods, "Should we subject this good to the calculus of social utility?" With moral goods, however, the situation is much different; the proper response to this sort of value is *not* to maximize it. So, while the bargainers may be motivated to promote certain natural goods by adopting a certain moral system, they might also be motivated to adopt a moral system that recognizes certain moral goods, *even if* adopting this particular moral system requires sacrificing the optimistic distribution of natural goods that might otherwise be available in a moral system that does *not* reflect these moral goods.

Let us suppose, as is plausible enough, that a set of rational bargainers will be motivated to do just this. Specifically, they will be motivated to accept a moral system that *gives proper expression to their moral status as persons*. The moral system they will be motivated to accept, I will argue, will of necessity be a moderate morality. It is because contractarianism, as opposed to the other available alternatives, provides such an easy way of justifying a system that expresses these moral goods (namely, by appealing to the motivations of the bargainers) that it seems the natural route to take when responding to Kagan's general challenge to moderate morality – despite Kagan's claim in his second challenge. I will explain more precisely what I mean by the claim that my bargainers desire to adopt a moral system that gives proper expression to their moral status as persons in my discussion of the bargainers' motivations below. Also, I will have more to say about the precise way in which such a contractarian account leads to options, and thus to genuinely moderate morality, in chapter four.

It is clear, then, what my contractarian account must accomplish if it is to be successful in meeting Kagan's specific challenges to the moderate from chapter four, and thus if it is to succeed as a response to Kagan's overall challenge to the moderate in *The Limits of Morality*. If it is to meet Kagan's first challenge, it must first of all be capable of grounding constraints and their principled exceptions in a common source. And secondly, my contractarian account must lead to a genuinely moderate morality – hence options – if it is to meet Kagan's second challenge. This I hope to accomplish with the contractarian view I am about to lay out.

CHAPTER TWO: A Contractarian Response to the Challenge

Most contractarian accounts share common notions about the conditions that must obtain at the bargaining table: the parties to the agreement are thought of as being supremely rational, motivated to maximize self-interest, and so forth.¹² I accept all these conditions as well (although I do not find it necessary to stipulate, with Rawls, that my bargainers are deliberating behind the ‘veil of ignorance’ – though I will not argue this point here). However, my account diverges from the standard contractarian stories in that it envisions the motivation of the bargainers’ to be twofold. This twofold motivation is in line with my earlier suggestion that we might expect a set of rational bargainers to be concerned with two types of goods. In other words, I fault the standard contractarian accounts for the same reason I faulted Kagan’s depiction of the contract approach above: while the standard accounts may adequately represent the bargainers’ attitudes towards natural goods, they do not sufficiently capture the attitude they can reasonably be expected to bear towards *moral* goods. And if we *do* consider the bargainers’ attitude towards moral goods, we will get a different story. In this chapter, I offer a rough outline of just such an account, which does take the bargainers’ attitude towards moral goods into account.

To begin with, I imagine my bargainers to be motivated to maximize self-interest, as are David Gauthier’s bargainers. But this motivation alone neglects to capture the bargainers’ attitude with respect to moral goods. And I imagine that the bargainers have a keen interest in at least this one moral good: they want the system of morality they adopt to be properly expressive and respectful of the *moral status* they bear as *persons*. Thus, the twofold motivation of my bargainers might be formulated as follows:

1. My bargainers are motivated to adopt a set of rules that, if generally followed, will be to everyone’s mutual welfare insofar as it seeks to maximize everyone’s self-interest, or at least seeks to maximize everyone’s ability to promote their own self-interest. Throughout this thesis, I will refer to this first motive as the ‘self-interest-maximizing motive.’

¹² For a look at some prominent recent contractarian accounts, see Gauthier, Scanlon, and Rawls.

2. My bargainers will be motivated to formulate a set of rules that – while of course seeking to promote self-interest – is nevertheless tempered by the need for the moral system properly to respect and to give expression to their moral status as persons. I will refer to this as the ‘personhood-respecting motive.’

The bargainers’ first motivation seems relatively unproblematic and in need of little explanation here. Again, it is intended to be the same sort of motivation that is appealed to in many of the traditional contractarian accounts. Two questions seem immediately to arise, though, regarding the bargainers’ *second* motivation: First of all, what is the ‘moral status’ that my bargainers believe themselves to bear insofar as they are persons? What is this status, in other words, such that my bargainers are motivated to recognize it by giving expression to it within the rules of the contract they adopt? And secondly, what exactly is it for a system of morality adequately to give expression to this status? Let us briefly examine the former question before going on, in the next two chapters, to a more detailed look at how exactly my bargainers will formulate a system of morality that gives proper expression to this status.

The moral status of persons that a set of rational bargainers will be motivated to recognize and respect in the moral system they agree to adopt has, I believe, at least three distinct, yet closely related, dimensions: persons are *worthy of concern*, they are *ends-in-themselves*, and they are *inviolable*. To say that persons are worthy of concern is roughly to say, more simply, that they *matter*. Persons matter because things matter to them. As such, there is always reason to promote the welfare of a person. The idea of a person as an end-in-itself is a somewhat richer concept, one that likely overlaps somewhat with both the ideas of persons as worthy of concern and as inviolable. For our purposes here, though, the distinctive feature of persons that we will want to capture by thinking of them as ends-in-themselves is the insight that, in Frances Kamm’s words, they “have a point even if they do not bring about greater good.”¹³ Following Thomas Nagel here we might say that, where persons are concerned, the agent-relative, first-person point of view matters at least as much as does the agent-neutral, third-person point of view.¹⁴ And

¹³ Kamm, 355.

¹⁴ Here I use Thomas Nagel’s terminology from *The View From Nowhere*: ‘agent-neutral values’ are ones that all persons equally have reason to promote. ‘Agent-relative values’, on the other hand, are those that have special reason-giving force only from the perspective of a few individuals, or one individual.

finally, while we might say that certain things should not be expected of persons, even for the sake of promoting the greater overall good, inasmuch as they are ends-in-themselves, so also certain things should not ever be *done* to them, even for the sake of promoting the greater overall good. This is just to say that persons are inviolable.

Thus far, I have suggested that a contract approach will adequately ground a moderate morality if we understand the bargainers' motivations, not as they are portrayed according to traditional contractarian accounts, but rather as twofold: as being concerned both with *natural* and *moral* goods. The remainder of this thesis will be devoted to explaining exactly how this twofold motivation leads to the constraints and options required for a successful reply to Kagan's challenge to moderate morality. Before we go on to do this, though, we should note that the two aspects of my bargainers' motivation do not amount to the same thing. To say that the bargainers are motivated to promote self-interest is not merely another way of saying they desire to formulate a set of rules that gives proper expression to their moral status as persons. Instead, these two motivations are quite distinct, and may at times even be in tension with each other. For instance, as we shall soon see, there may be some actions that are perfectly compatible with restrictions on behavior that would arise from the first motive alone, but that are incompatible with the bargainers' *second* motive. What this means is that in our effort to ground constraints, their principled exceptions, and options all in the common source of our bargainers' desires, we must take extra care that the constraints, exceptions and options we formulate be such that they adequately fit *both* dimensions of their twofold motivation.

CHAPTER THREE: Modified Constraints

In this chapter, I will discuss exactly what sort of constraints will arise from the contractarian account as I have laid it out thus far. Certain aspects of my bargainers' motivations lead more naturally to a grounding of constraints than do others; therefore, I will be focusing in this chapter on the bargainers' self-interest-maximizing motive and their desire that their moral system properly recognize persons' *inviolability*.

I will argue that from the bargainers' motivation to maximize self-interest we get the typical restrictions on harming, killing, stealing, lying and so forth, on the grounds that it is in everyone's mutual self-interest to refrain from engaging in these sorts of activities. Similar restrictions, I shall argue later, will also arise out of the bargainers' second motive – though this motive will result in the bargainers' adopting genuine deontological constraints, rather than mere restrictions.¹⁵

Next, I will go on to note that adopting these restrictions or constraints involves a certain trade-off in each case. The bargainers will be motivated to make these trade-offs as slight as possible; consequently, then, they will adopt *modified* constraints. As we shall see, this results in the constraints' being 'defeasible,' as the adoption of absolute restrictions or constraints would oftentimes run afoul of the bargainers' initial motivation to adopt the social contract in the first place. To illustrate this, I will point to the bargainers' need to minimize the trade-offs threatened by free-riders and note that the very possibility of such a threat illustrates the fact that *any* contractarian account, if it is to have any hope of succeeding, must allow for exceptions. We will examine the way in which the defeasibility of the constraints emerges both from the bargainers' self-interest-maximizing motive and from the concern with persons' inviolability that constitutes part of their personhood-respecting motive. With regard to the former motivation, we will note that the restrictions were only erected in the first place because they worked to everyone's mutual advantage, while free-riders seek to exploit these very restrictions and use them to everybody's mutual *disadvantage*. In this case, the restrictions are no longer

¹⁵ As I am using the terms here, 'restrictions' are prohibitions on certain types of behavior adopted because of their utility, which can thus be overridden if utility so dictates. 'Constraints,' meanwhile, are prohibitions on certain types of behavior grounded in some source other than utility; though it may be

serving their original purpose; in fact, they are acting directly *contrary* to their original purpose. Thus, the bargainers' first motive will lead them to adopt a sort of 'defeasibility clause' to complement the restrictions they've already adopted, and to address just this concern. Such a clause will allow them to override the restrictions anytime doing so is a necessary means to avoiding the dangers uniquely posed by free-riders.

As it emerges from the bargainers' first motive, though, this 'defeasibility clause' would appear to permit intuitively impermissible exceptions. However, the clause takes on additional features when it is modified so as also to fit the bargainers' second motivation. The personhood-respecting motive legislates against seemingly impermissible exceptions (for example, the exception to the restriction on killing others required in the case of 'Transplant.'¹⁶) while still allowing for exceptions in cases of, for example, self-defense. The result of all this, then, is that we see that the bargainers' twofold motivation will lead them to adopt genuinely deontological *constraints* and a 'defeasibility clause' that generates the principled exceptions that would seem to be an integral part of any worthwhile system of morality. I will then conclude the chapter with a brief discussion of *thresholds*, which constitute a further modification to the constraints.

I. The self-interest-maximizing motive and restrictions on the pursuit of individual good

Let us start by examining the bargainers' first motivation – their desire to maximize self-interest. It has long been the hallmark of social contract theories that they demonstrate that a set of rational bargainers will accept a set of advantage-overriding restrictions on their behavior because, ultimately, accepting these restrictions is in their best self-interest. As David Gauthier puts the point, "Duty overrides advantage, but the acceptance of duty is truly advantageous."¹⁷ This feature of contractarian thought is fairly obvious and standard, and can easily be illustrated by the example of the "Prisoners' Dilemma."

possible that these prohibitions too may be overridden (in threshold circumstances, for example), they cannot be overridden just anytime doing so will result in the greater good.

¹⁶ 'Transplant': the example comes from Phillipa Foot, and asks us to imagine a scenario wherein a relatively healthy, innocent hospital patient is sacrificed so that his organs may be re-distributed to save the lives of five otherwise terminally ill patients.

¹⁷ Gauthier, David. *Morals By Agreement*. New York: Clarendon Press, Oxford, 1986. p. 2.

A. The Prisoners' Dilemma

This common illustration, borrowed from game theory, works as follows: imagine that you and a friend are apprehended by the authorities and charged with a crime. You are each taken to two separate interrogation rooms, from which you cannot communicate with each other, and are presented with two options: you can either *keep mum* in response to the authorities' accusation, or you can *betray* your friend by implicating him as the sole perpetrator of the crime. You and your friend are also advised that one of the following three scenarios will obtain, depending on how each of you responds: (1) If both you and your friend keep mum, you each spend three nights in jail. (2) If one friend keeps mum while the other betrays, the friend who betrays goes scot-free, while the silent-but-betrayed friend spends a month in jail. (3) If both friends betray each other, each of them spends a week in jail. Obviously, each friend's best bet is to betray the other party, because regardless of the other party's decision, betraying your friend offers you the best odds of receiving the least amount of punishment. But, of course, if both parties betray each other, they each needlessly wind up worse off than they could otherwise have been. For had they been able to communicate with each other ahead of time, and had they been able to trust each other, they each could have agreed to keep mum, thus ensuring that both parties would have wound up better off than they would have had they each betrayed the other – they each would have spent a mere three nights in jail, rather than a week.

The situation in standard contractarian or social-contract accounts works much the same way: in a lawless society, or in the state of nature, everything is permitted. Thus, if I know that there is nothing stopping my neighbors from lying to me, stealing from me, harming me, or even killing me for the sake of their own personal gain, then I would be foolish if I did not also try to maximize my own self-interest by lying and stealing from all my neighbors. Under such an arrangement, though, everybody winds up worse off than they need be. For while my advantage is promoted by the fact that I can violate my neighbors all I want in order to advance my lot in life, this benefit is more than outweighed by the fact that all my neighbors are themselves each trying to steal from me, lie to me, or harm me . . . anything that might advance *their* lot in life. The result is that I am much worse off, because anything I might happen to gain by violating my neighbors

is balanced out by, or likely outweighed by, everything I most certainly will lose to all my advantage-seeking neighbors. And this same plight of course goes for *everybody*. Being rational, then, we would all agree *not* to engage in stealing, lying, harming, killing, and so forth, provided everyone else refrains from these activities as well. It is the contractarian's contention, then, that moral rules derive their normative force from the fact that they are the ones we would agree to in such a situation. Such, then, is the contractarian account of the origin of the standard restrictions against harming, lying, and stealing. These restrictions, we have just seen, can be founded in my bargainers' self-interest-maximizing motive alone.

B. The trade-off

In addition to the restrictions considered above, though, the bargainers' first motive will lead them to adopt certain sorts of principled exceptions to these restrictions. For consider this interesting feature of contractarian scenarios, which we have already considered briefly: the bargainers in the social-contract setting are negotiating with each other to see which restrictions they will adopt and adhere to as a preferable alternative to living in the state of nature, a condition which we have already seen to be detrimental to all parties. The restrictions that will be agreed upon, then, are adapted and suited to conditions encountered in the state of nature. It is inevitable, however, that these conditions will change somewhat once the community enters a state of society. Obviously, one of the new conditions that the bargainers can expect to encounter in this new, more civilized setting is that they can expect a change in the motives of their fellows. No longer will their neighbors be the purely-self-interest-maximizing agents that they were in the pre-social-contract condition; rather, they will be agents who, though still self-interested, will pursue their own well-being under the restraints imposed upon them by the restrictions to which they've agreed. However, it seems almost as obvious to expect that in the new civilized state, they can expect something else as well: a host of free-riders who will recognize the new opportunities that such a state of civilization affords, and will seek to profit at others' expense by ignoring the restrictions on advantage to which they had agreed, all the while expecting others to remain committed to them. Such a situation, it is important to note, can only arise once a community has agreed mutually to adhere to a set of advantage-overriding constraints; these free-riders

cannot pose a threat to our well-being, or to the overall good, in the state of nature. It is a community's very acceptance of a set of agreed-upon restrictions *itself*, and nothing else, which opens up the possibility of such cases. Obviously, then, the adoption of certain agreed-upon restrictions brings with it – due to the very nature of the community that will arise should these new rules be implemented – the requirement that certain exceptions be built into the structure of the restrictions. Thus, we see that any such account permits – indeed, *requires* – the possibility of principled exceptions to at least certain restrictions in order for the entire apparatus to be effective. And these exceptions are grounded in the same principle that grounded the restrictions in the first place – in this case, the bargainers' self-interest-maximizing motive. Without these exceptions, as we noted above, the set of restrictions will serve merely as one elaborate mechanism to be exploited by, and serve the interests of, violators and free-riders.

The defeasibility of these restrictions will take two forms. First of all, we will permit sanctions for the violators, as a means of punishment for the sake of deterrence and social protection. This feature of contractarian accounts – that they allow for exceptions to the rules for the sake of punishing violators – is widely recognized. But furthermore, we must also build in reasonable exception clauses, allowing (for example) for proportional harm to attackers in self-defense.

C. Minimizing the trade-off: the 'principle of defeasibility'

So far, then, we have established that a certain feature of contractarianism – namely, the threat of free-riders that arises once a community adopts the contract – yields the necessity of allowing for certain principled exceptions to the restrictions that make up the contract – the kinds of exceptions required both in the case of sanctions for wrongdoers and in some cases of self-defense. But which exceptions can be allowed, and which cannot? How do we determine which exceptions, in other words, are 'principled'? Do the bargainers need to enumerate, for each restriction on behavior they agree to adopt, a list of scenarios in which the restriction may be defeated or overridden, as part of the terms of the contract? This seems as burdensome as it is *ad hoc*. Instead, I want to argue that in addition to agreeing to whichever restrictions they happen to adopt, the bargainers need simply agree to one further principle that will generate all the exceptions they need, rather than specifying each individual acceptable exception. This principle might be

called the contractarian ‘principle of defeasibility’, and would seem to be an essential component of any successful contractarian account. The motivation for it has already been discussed: the implementation of the social contract will itself open up the possibility of a new kind of danger that the bargainers never confronted in the state of nature: the threat of free-riders. Hence, in their formulation of the contract, the bargainers will have reason not only to adopt principles that will minimize the evils they encounter in the state of nature, but they will also have reason to adopt a further clause that would seek to minimize the dangers uniquely presented by the implementation of the contract. Roughly stated, the ‘defeasibility principle’ would sanction exceptions to restrictions precisely insofar as those same restrictions are being exploited by free-riders seeking to further their own advantage at the expense of the rest of the constraint-abiding population. To the extent that a free-rider is breaking a restriction and violating an innocent person, a restriction may be overridden if doing so is necessary for the protection of the innocent, or to prevent the free-rider from succeeding in his efforts to profit at the expense of the moral system. One might put the same point by saying that “protection under the contract is conditional on conformity with the contract,”¹⁸ and that the extent to which someone breaks the contract in pursuit of personal gain is the extent to which he forfeits his rights against others that they continue to adhere to the rules of the contract when dealing with him.

Self-defense, of course, provides a concrete illustration of this: restrictions against harming others are agreed to because, in the long run, they serve everybody’s greater good. However, a free-rider should not expect his fellows to remain committed to these restrictions while he breaks them in an effort to achieve undue gain; the contractarian principle of defeasibility will permit the intended victims of the free-rider’s offense to override the restriction if doing so is a necessary means to protecting themselves from the free-rider’s violations.

This, then, is the way restrictions and their principled exceptions arise out of the bargainers’ self-interest-maximizing motive. But our task is not done. For as we noted above, we must be sure that whatever rules our bargainers agree to fit both dimensions of

¹⁸ Kagan 1989, p. 135

their twofold motivation. Let us go on to consider the shape of the modified restrictions that arise out of our bargainers' second, personhood-respecting motive.

II. The personhood-respecting motive and restrictions on the pursuit of individual good

My bargainers' second motivation will of course also lead them to adopt some sorts of restrictions. At first glance, these restrictions will have pretty much the same shape as the restrictions that emerge from the self-interest-maximizing motive. And the reason for this should be fairly clear: lying to, stealing from, and harming other persons pretty obviously constitutes disrespecting them and disregarding the moral status they bear insofar as they are persons. Any moral system that will give proper expression to this status, then, must surely prohibit such behavior. Upon a closer examination, though, we will see that the restrictions arising out of the bargainers' second motive have additional features that are not shared with the simple restrictions that emerge from the bargainers' first motive. For the restrictions that result from the bargainers' second motive will not merely be rules adopted because they are expedient in promoting everyone's greater self-interest. Instead, these restrictions will be adopted in order to give expression to the moral status of persons. As such, they will confer *rights* upon all parties to the contract. Assigning rights to the members of the moral community is not simply a matter of expediency. It is a matter of determining what may or may not be done to or required of a person, *regardless* of the promotion of overall self-interest that could thereby be enacted. While the assigning and honoring of rights more often than not *is* expedient in bringing about everybody's greater self-interest, a system of morality that features genuine constraints and rights demands that these constraints and rights be honored, even on those relatively few occasions when *not* honoring them would be more expedient. So the bargainers' personhood-respecting motive leads not simply to restrictions on harming, lying, stealing, and so forth, but to genuine deontological constraints against these things. On their face, these constraints bear much the same shape as the mere restrictions that the bargainers' arrived at out of their first motive; the distinctive features of genuinely deontological *constraints* will become clear, though, when we consider the extent to which these constraints are *defeasible* – that is, the extent to which they will admit of principled exceptions. In what follows, I will examine the way in which the bargainers' particular motive to formulate a set of rules that recognizes and preserves persons'

inviolable status alters the shape of the restrictions and their exceptions as they arise solely out of the bargainers' self-interest-maximizing motive.

To say that my bargainers "wish to formulate a set of rules that gives proper expression to the inviolable status they bear as persons" requires a little bit more explanation. To see what exactly is meant by this claim, let us examine in closer detail the notion of *inviolability* and see what would be involved in adopting a set of rules that gives proper expression to it. There are two main ideas that need attention in my stipulation that my bargainers desire the system of morality they adopt to respect their inviolable status as persons: first, the conception of 'inviolability,' and second, the notion of an ethical system's 'respecting' or 'giving proper expression to' this inviolable status.

A. Inviolability

Much has been made, in the philosophical discussion of recent years, of the ideas that persons are in some sense 'inviolable' and that it is a desired feature of moral systems that they 'respect' or 'give expression to' this inviolability.¹⁹ It is generally thought that the intuitions driving this discussion (and, presumably, my bargainers) are non-consequentialist in nature. Nevertheless, some consequentialists have argued that these intuitions regarding the sanctity of the individual and the inherent evil of violating persons can be incorporated into a pluralistic theory of the good.²⁰ As such, consequentialist moralities might be able to incorporate these intuitions into their systems by assigning great negative weight to such violations and sanctioning actions or policies that seek to minimize such transgressions. It is important to see that such a point of view does not adequately capture the insight driving the discussion conducted by Kamm, Nagel, and Quinn, nor does it fully reflect the intuitions my bargainers seek to see recognized. This is especially important for our purposes, as we are committed to demonstrating that our contractarian account leads not to consequentialism, but to genuine, *deontological* constraints. If such a concern on the part of the bargainers might be shown just as easily to lead to a consequentialist as to a non-consequentialist morality, then our hope of grounding a moderate morality in the social-contract setting will fail.

¹⁹ See, for example, Quinn (1989), Kamm (1992) and Nagel (1995).

²⁰ See, for example, Amartya Sen, "Rights and Agency," *Philosophy and Public Affairs* 11 (1982), pp. 3-39.

The reason consequentialist endeavors to 'mimic' deontological constraints -- even personhood-respecting, violations-minimizing constraints -- are bound to fail is that they fail to recognize the very conception of the moral status of a person that Kamm, Nagel, and my bargainers wish to respect in the first place. Frances Kamm clearly recognizes the distinction between these two conceptions of the moral status of a person in her discussion of the subject, when she says that "giving greater weight to the negative of harming than to the positive of benefiting, I believe, represents the priority morality gives to the inviolability of the person over his status as recipient of such benefits as length of life."²¹ Warren S. Quinn makes a similar recognition when he writes that morality, in granting to each person primary say over what may be done with his body, mind, and life (rather than giving primary control of these things to the community, as items to be exploited in the service of its greater overall good), "recognizes [each person's] existence as an individual with ends of his own -- an independent *being* . . . Were morality to withhold [this recognition] . . . it would picture him not as a being in his own right but as a cell in the collective whole."²²

Both these authors seem to have clearly in view two competing notions of a person and his or her status in a moral system: on the one hand, there is the idea of a person primarily as both a recipient and a source of both benefits and harms, and as such, a means to the maximization of the former and the minimization of the latter. On the other hand, there is Quinn's characterization of a person as a *being* ("the ordinary sense of "being," in which human persons, gods, angels, and probably the higher animals -- but not plants, cells, rocks, computers, etc. -- count as beings."²³), or Kamm's recognition of persons, not as "mere means to the end of the best state of affairs, but ends-in-themselves, having a point even if they do not serve the best consequences."²⁴ No doubt we wish to affirm that human beings bear both these statuses in some degree or another; the question becomes whether or not one status is the primary one and takes precedence over the other, and if so, which one.

²¹ Kamm 1992, 382.

²² Quinn, 156.

²³ Quinn, 161.

²⁴ Kamm 1992, 358-9.

It seems pretty clear that from the consequentialist point of view, persons bear the former status with much more force than they bear the latter. In fact, the extent to which persons bear the latter kind of status merely serves to establish their recognition as ends-in-themselves, or their inviolability, or anything else that is attendant with that status -- indeed, that very status *itself* -- as a good or a benefit, in the sense in which persons are primarily the recipients of benefits, which can be reckoned in along with the other ordinary benefits (goods of life, health, pleasure, etc.) of which they might be thought to be recipients. In other words, this 'inviolable' status is merely a *natural* good, subject to the utilitarian calculus which seeks the optimific distribution of goods among its subjects. Only in such a way could a consequentialist argue that violations of personhood are a very bad thing, which should be minimized even if that requires violating a few to prevent a greater number from being violated.

From a non-consequentialist point of view, though, the inviolable status of persons is a *moral* good. If persons are conceived of primarily as inviolable ends-in-themselves in the way that Quinn and Kamm have characterized them, things look very differently than they did on the consequentialist account. If persons do, as Quinn thinks, properly *own* their bodies, minds, and lives, then they are *not* subject to the calculus of social utility in the way many consequentialists think they are. For, "whether we are speaking of ownership or more fundamental forms of possession, something is, morally speaking, his only if his say over what may be done to it (and thereby to him) can override the greater needs of others."²⁵ To say that a person is inviolable is just to say that his or her life, body, mind, and so forth are his or hers to do with what he or she will, within the constraints of morality – constraints, it should be noted, which exist in the first place for the primary purpose of recognizing the fact of persons' inviolability. If persons are inviolable, then, they may not be harmed, lied to, or otherwise wronged, and furthermore, their lives may not be subjected to the utilitarian calculus – not for the purposes of increasing overall welfare, and not even for the purposes of ensuring that as few persons as possible are violated. It is this latter concept of the status of persons that my bargainers are seeking to see respected and recognized within the set of rules they agree to formulate. And as we are already starting to see, not just any selection of moral

²⁵ Quinn, 156.

principles will do the trick here. For on such a view, the inviolability of persons is not a natural good to be promoted *within* a moral system, but a *moral* good to be recognized *by* the system.

B. Morality's 'giving proper expression to' the inviolable status of persons: genuinely deontological constraints.

By arguing that persons are primarily bearers of the first type of status, Kamm, Quinn, Nagel, and my bargainers are insisting that a certain inviolability accompanies personhood. Such inviolability rules out the possibility that persons' lives might be sacrificed for the greater good for any reason -- even if that reason is the reduction of similar violations to other persons -- and therefore calls for the erection of genuinely deontological constraints for the sake of recognizing this inviolability. Thus, while the 'person as recipient of goods' conceptions fits in best with consequentialism (indeed, we might even say that each one implies the other), the 'person as inviolable' conception fits in best with non-consequentialism. In fact, the inviolable status of persons -- though it is still compatible with thresholds²⁶ -- is uncompromising in its demand for genuine deontological constraints against, say, harming others.

The bargainers, then, will be motivated to adopt a moral system that reflects their inviolable status -- they will adopt a system of *deontological* constraints. Notice that this status is the sort of thing that must be *reflected* by the rules of morality -- it is not something that should or even *can* be 'brought about' (or 'promoted', or 'maximized') by the ethical system. In fact, it is the very essence of a properly 'deontological' constraint that it be non-teleological. It is a matter of what *can* or *may* (justifiably) happen to a person, not what actually *does* happen to him. Furthermore, the erection of these constraints results in the parties to the contract being endowed with certain rights. These points have been stressed heavily in the recent discussion of this topic. Warren Quinn states that,

The value that lies at the heart of my argument -- the appropriateness of morality's recognizing us as independent beings -- is in the first instance a virtue of the moral design *itself*. The fittingness of this recognition is not a goal of action, and therefore not something that we could be tempted to serve by

²⁶ 'Thresholds' are points above which the evil that would result from adhering to a constraint might be so great as to override the constraint. They will be discussed in more detail in section III-E below.

violating or infringing anybody's rights.²⁷

Thomas Nagel carries out an extensive discussion of the relationship between these constraints, rights, and inviolability in Part III of his article "Personal Rights and Public Space." He argues that, in order fully to understand the role of rights as a part of a non-consequentialist morality, we need to understand "the status conferred on all human beings by the *design* of a morality which includes agent-relative constraints of this kind. The status is that of a certain kind of *inviolability*, which we identify with the possession of rights."²⁸ Rights belong properly to a non-consequentialist, moderate morality precisely because they

prohibit us from *doing* certain things to *anyone* but do not require that we count it equally a reason for action that it will prevent those same sorts of things *being done* to someone else, but not by oneself . . . Rights tell us in the first instance what not to *do* to other people, rather than what to *prevent from happening to them*.²⁹

This, to Nagel, leads to a certain "paradoxical" feature of constraints and rights, insofar as a right or a constraint "may sometimes forbid us to do something that would minimize its violation -- as when we are forbidden to kill one innocent person even to prevent two other innocents from being killed."³⁰ But this paradoxical feature just arises from the fact that, not only are the goods promoted by a set of moral rules important, but it is *also* important that a certain status be conferred or recognized by such a system:

If . . . it is permissible to kill the one to save the two [from being killed unjustly], that implies a profound difference in the status of everyone . . . In the world with no rights and fewer killings, *no one* would be inviolable in a way in which, in the world with more rights and more killings, everyone would be -- including the victims.³¹

For

what actually happens to us is not the only thing we care about: What *may* be done to us is also important, quite apart from whether or not it *is* done to us . . . we are trying to explain the moral significance of agent-relative rights by say-

²⁷ Quinn, 157-158.

²⁸ Nagel, 89.

²⁹ Nagel, 88.

³⁰ Nagel, 90.

³¹ *Ibid.*

ing that not only is it an evil for a person to be harmed in certain ways, but for it to be *permissible* to harm the person in those ways is an additional and independent evil.³²

As this applies to our contractarian view, we might say, “Natural goods are not the only concern of a set of rational bargainers; they will be concerned with moral goods as well.”

Frances Kamm perhaps puts the point against the consequentialist best of all, when she asks,

Would it not express greater concern for this conception of the person [as an inviolable end-in-itself] if we minimized the number of occasions on which people were not permitted to choose, even if this meant occasionally depriving someone of choice by obliging him to make a big sacrifice that ensured that others might choose? My view is that *permitting* this means to minimization . . . would defeat the very ideal of the person as an end-in-itself that was supposedly the object of concern. For if such an obligation were appropriate, the individual would no longer be someone who was not 'for' the greater good, as he would be available for minimizing interference with the value of choice.³³

C. The trade-off

But Kamm also takes note of another almost 'paradoxical' element of the right/constraint structure of such morality, similar to the paradoxical element that Nagel wrote about. This involves the trade-off required, which both Kamm and Nagel discuss: we may foresee that there actually may be *more* rights- or personhood-violations (more killings, stealings, etc.) in this system that supposedly gives such a high premium to constraints against violating persons, than we would in certain consequentialist systems that sanctioned any means necessary to the minimization of such violations. However, as Kamm points out, by adopting such a morality, we are not actually sanctioning these violations, even as we foresee that they may occur. She writes:

If morality permitted minimizing violations of persons by violating other persons, then each of those saved as well as those persons used to save others would be less inviolable. It is the permission, not any actual violation of persons, that makes this so. If more violations of constraints actually occur because violations are not permitted, this does not mean that morality endorses the correctness of these harmings. More

³² Nagel, 91.

³³ Kamm 1992, 359.

people are harmed, and so the chances of each of us being harmed may well be greater; but the conception of each person that is morally endorsed involves a high degree of inviolability. We may all lead harder lives, but our dignity is greater. We may actually prefer this trade-off.³⁴

While Kamm may be right that we would prefer this trade-off, I think it is once again obvious that we would desire as slight a trade-off as possible. And this just goes to show the importance of having *sanctions* in our moral system: not only must we be permitted to perform acts, normally prohibited, against offenders after the fact of their having violated the rights of others (for instance, we should be allowed to imprison them as a means of disciplining them), but we must also permit exceptions to constraints as a means of protecting ourselves from attacks on our inviolability right as they occur. Thus, this 'paradoxical' feature of our contract-based, personhood-respecting moderate morality makes all the more pressing the need for our successful effort to ground principled exceptions to constraints in cases of self-defense. For, if constraints could never be defeated, and inviolability absolute, in the sort of way that it could never be forfeited, imagine what life would be like. We would be totally at the mercy of people who sought to exploit these constraints for their own purposes, seeking to profit from their *own* violations of constraints, secure in the knowledge that we ourselves remained committed to honoring the same constraints, even at the expense of allowing a greater number of rights violations. While this may be acceptable in some instances -- we do not think morality should require, or even permit, our acting in accord with the terrorist who demands we kill one innocent person, lest he murder two -- nevertheless, we do not think that morality should legislate against our using force and inflicting harm, if necessary, on someone who is himself attempting to violate our rights by harming us. So we must allow that, in some cases at least, violators of inviolable persons thereby relinquish or forfeit their inviolable status; otherwise, the set of constraints will merely serve as one elaborate mechanism to be exploited by, and serve the interests of, violators and free-riders.

³⁴ Kamm 1992, 383.

D. Minimizing the trade-off

So the bargainers' second motive will lead to their acceptance of some sort of defeasibility clause as well. However, it will have a different shape than the clause that arises solely from the bargainers' first motive. This is so because many of the exceptions that would be permitted by the defeasibility clause as it emerges for the bargainers' first motivation alone would require disrespecting the inviolable status of persons that the bargainers are motivated to respect. For example, the exception to the restriction on killing others required in the case of 'Transplant' might be permitted by the self-interest-maximizing motive. This is so because permitting such exceptions would increase the average overall welfare; any bargainers seeking *only* to maximize their welfare would surely be rational to agree to exceptions that will increase their chances of being better off. However, my bargainers are *not* interested only in maximizing their chances of being well off, and the bargainers' personhood-respecting motive will require the modification of this defeasibility clause. As we saw in the discussion of inviolability above, to say that a person is inviolable is in part to say that he or she cannot be required to sacrifice his or her well-being simply because doing so would serve the greater overall good. Shaping the defeasibility clause so that it matches the bargainers' second motivation, then, eliminates the possibility that it will sanction 'Transplant'-style exceptions.

Even though the bargainers' second motive does allow *these* types of exceptions, however, it will still press upon them the necessity of sanctioning *some* exceptions, and thus of adopting some sort of defeasibility clause. It is just that the exceptions allowed cannot compromise persons' inviolable status. In order to make the trade-off Kamm discusses as slight as possible, the bargainers will be motivated to stipulate that, although rights-violations are not permitted even to minimize the occurrence of rights-violations, those who *do* violate others' rights forfeit their own rights to the same degree. Thus, while harming persons disrespects the moral status they bear as persons, and as such is prohibited by the constraints the bargainers' adopted due to their personhood-respecting-motive, someone who himself is harming or has harmed innocent people has forfeited at least some of his right against us that we not harm him. Therefore, by harming him in self-defense, or by harming him by meting out sanctions against him, we are not violating

him or disrespecting his moral status as a person, because we are not actually infringing any of his rights.

E. A further modification of the constraint: thresholds

We must note briefly that one other feature of the constraints emerges from the bargainers' twofold motivation: thresholds. The belief in thresholds – that is, the belief that the evil that will result from *not* breaking a constraint or violating a right might sometimes be so great that the constraints can in those cases be overridden – emerges also from the bargainers' second motive. We have already seen how this personhood-respecting motive leads to genuinely deontological constraints. This was because in their concern to formulate a set of rules that gave proper expression to the moral status of persons, the bargainers sought properly to recognize persons' *inviolability* with their rules. Hence they adopted *constraints*, rather than the mere *restrictions* that emerged from the bargainers' self-interest-maximizing motive. However, we must recall that *inviolability* is not the only feature of this moral status of persons that the bargainers wished to respect. According to this view of persons' moral status, people are also 'worthy of concern.' Although this feature of persons is generally not as strong as their *inviolability*, it does count. And indeed, there may be cases in which the 'worthiness of concern' of the many might be strong enough to override the *inviolability* of the one. It is important to note, though, that in agreeing to such thresholds, the bargainers are still in keeping with their personhood-respecting motive, in exactly the same way in which agreeing to the *inviolability-honoring* constraints in the first place represented their personhood-respecting motive. It is simply that the bargainers recognize that, though important, the *inviolability* of persons is not the only dimension of their moral status that they wish their moral system to respect – hence the incorporation of thresholds into this moral system.

III. Conclusion: The shape of the modified constraints.

In closing, then, we see that the bargainers' self-interest-maximizing motivation will result in their adopting particular restrictions against lying, stealing, harming and killing, as the overall adherence to these restrictions will increase everyone's average

welfare. The bargainers' personhood-respecting motive then adds further content to these restrictions by insisting that they be genuine deontological constraints. This distinction between 'restrictions' and 'constraints' is most fully illustrated by considering the different ways in which they are defeasible: restrictions can be overridden for the sake of the greater good (for they were erected in the first place solely for the purpose of serving the greater overall good), while constraints cannot. The defeasibility of restrictions is compatible with the action required in 'Transplant'; the defeasibility of constraints is not.

These constraints express more than just the bargainers' desire to maximize overall well-being – they give voice to the bargainers' desire that their moral system give adequate expression to their moral status as persons. The erection of constraints respects the sanctity of each person, as they protect persons from infringement not only insofar as this protection leads to the greater overall good, but even in those cases where infringing persons actually would be optimific.

Still, though, these constraints admit of principled exceptions. In fact, special features of the social-contract scenario seem to *require* the adoption of some sort of 'defeasibility clause.' Such a clause generates the exceptions necessary to prevent the social contract, once implemented, from collapsing back onto itself, as it were, by enabling free-riders to exploit the system to their own advantage. The bargainers' first motive leads to the formulation of just such a principle, which justifies exceptions anytime adherence to the restrictions would be directly contrary to the initial self-interest-maximizing motive that led to the adoption of the restrictions in the first place. Once again, though, the bargainers' second motive steps in to modify this defeasibility principle, requiring that only exceptions that are compatible with the moral status of persons may be permitted. (This is just to say that only the exceptions that are compatible with genuine *constraints* may be permitted.) The result so far, then, is a system of morality which yields deontological constraints that are tempered by thresholds and admit of exceptions in cases of, for example, self-defense and sanctions for wrongdoers, but do not admit of just any exception that will increase the overall good. Furthermore, these constraints and their exceptions are grounded in a common source: the bargainers' twofold motivation.

CHAPTER FOUR: Options

So far, we have seen how our contractarian story grounds constraints and their principled exceptions mainly in the bargainers' self-interest-maximizing motive and in the inviolability-respecting aspect of their personhood-respecting motive. After briefly considering how they are compatible with these same two elements of the bargainers' motivation, this chapter will proceed to give a more detailed explanation of how exactly *options* might plausibly be generated from the two aspects of the bargainers' second motive that have heretofore been little-discussed: the concerns to express persons' *worthiness of concern* and their status as *ends-in-themselves*.

I. Self-interest and options: the appeal to cost

People often try to defend options by appealing to the 'cost' imposed on persons if they had always and only to fulfill their duty to promote the overall good at the expense of pursuing individual goals.³⁵ This is a cost, it is argued, that morality cannot reasonably expect persons to bear. Such an argument clearly rests on an appeal to self-interest, and thus seems to be the sort of justification of options that might arise solely from the bargainers' first motive.

While perhaps initially plausible, this argument suffers the drawback that it seems to be incompatible with our previous justification of constraints. For if certain actions promoting the greater overall good may legitimately be avoided on grounds that their cost to the agent would be too high, might not adherence to certain constraints *also* be avoided in some circumstances, if such adherence would also be of great cost to the agent? That is, if we can appeal to the excessive cost demanded of us as a reason not to give significant portions of our earnings away to charity, why might we not also make this appeal to cost as a reason not to refrain from secretly murdering our rich uncle Albert, from whom we stand to inherit a small mint?³⁶ Adherence to the constraint against killing others, in this case, is costing us just as much as, or even more than, many actions that would promote the overall good, but which we may think we legitimately have the option *not* to perform. Such an appeal to cost, then – at least without further

³⁵ Kagan discusses this idea on pp. 21-4 and in Chapter 7 (pp. 231-270) of *The Limits of Morality*.

³⁶ This example is discussed by Kagan, p. 4.

consideration – cannot serve as our exclusive justification for options. But again, the self-interest-maximizing motive is *not* the only motive of which our bargainers will want to avail themselves. While the appeal to cost may be a partial but nonetheless legitimate component of our defense of options, let us see if a more successful and thorough grounding of this last component of moderate morality is available to us in the bargainers’ second motivation.

II. The personhood-respecting motive and options

We may be tempted to ask ourselves first of all if options are even the sort of thing we need to justify in arguing for our moderate morality on the basis of the bargainers’ twofold motivation. For it may be objected that ‘options’ presuppose that there is something that needs to be opted out of in the first place. For Kagan, at least, there *is* a *pro tanto* reason to promote the overall good – a reason, he thinks, the moderate is committed to accepting.³⁷ And so on his account, it is precisely the moderate’s acceptance of the *pro tanto* reason that necessitates her defense of options as part of her successful overall defense of moderate morality. *But*, we may think, we have been arguing so far on the basis of a particular conception of the bargainers’ motivation, and as we have conceived it, this motivation – particularly, the bargainers’ motivation to capture within the rules of the contract a certain ideal of the *moral status of a person* – is itself incompatible with the existence of the *pro tanto* reason to promote the greater overall good. This is so because our conception of the moral status of a person precludes persons’ being in any way merely instruments of serving the greater overall good. But the *pro tanto* reason to promote the overall good presupposes a picture of persons as primarily or exclusively recipients of benefits and services and sufferers of evils, and also as means thereby of maximizing the former and minimizing the latter. Thus, the argument might continue, our very defense of moderate morality has already, from the start, eliminated even the need to defend options; we will already have succeeded in defending moderate morality against the extremist, then, simply by grounding legitimate *constraints* in this conception of the moral status of a person.

But while it is true that our bargainers’ desire to preserve persons’ status as ends-in-themselves does preclude persons’ being conceived *merely* as instruments of the

³⁷ See his arguments on pp. 48-52.

greater good (and, of course, partially constitutive *of* this greater good), it does not deny entirely that people *do* have duties towards this end. For in addition to arguing that persons are inviolable and ends-in-themselves, we have also stipulated (in the third aspect of our bargainers' conception of the moral status of a person) that persons are *worthy of concern*. And to say that persons are worthy of concern is precisely to say that they are the sorts of beings capable of generating a valid *pro tanto* reason to promote their good (whether individually, in acts of charity or supererogation, or insofar as their good is bound up with the good of the greater whole). As moderates, then, we are committed to this *pro tanto* reason after all; still, though, our conception of persons' moral status as a whole supports both the acceptance of this reason *and* options.

As Kagan is thinking of it, the *pro tanto* reason to promote the overall good is an *insistent* reason for action – one “demanding action in the absence of countervailing considerations.”³⁸ Such a reason *may* encounter countervailing reasons in the form of certain constraints on what may be done to promote the overall good (though Kagan is skeptical of this possibility), but as he argues at great length throughout his book, none of the reasons that might ever serve to override the insistent reason to promote the good will ever take the form of anything like *options*. Hence, he thinks, acceptance of the *pro tanto* reason commits one to extremism – and, short of one's ability to defend constraints, to consequentialism. But – though our affirmation that persons are indeed worthy of concern does commit us to accepting the *pro tanto* reason – does it require that this reason be an insistent reason? Might not the ‘persons as worthy of concern’ dimension of our personhood-respecting motive simply generate a non-insistent reason to promote the greater overall good? Such a non-insistent reason is one that, at any time an agent wanted to, could be invoked to justify her action (provided that that action does not violate a constraint), but is not in need of countervailing considerations in order for the agent to be justified in not acting on it.

I believe that in fact, the ‘persons as worthy of concern’ dimension of the bargainers' second motivation can and *must* generate a *non-insistent* reason to promote the overall good. For otherwise, this aspect of the personhood-respecting motive would be at odds with the belief that persons are ends-in-themselves – that they “have a point

³⁸ Kamm 1992, p. 356.

even if they do not bring about greater good.”³⁹ While the bargainers’ belief that persons are worthy of aid and concern generates an agent-neutral *pro tanto* reason, in general, to promote their well-being, their belief that persons are *also* ends-in-themselves generates a sort of agent-relative *pro tanto* reason for persons to promote their own individual good. If both of these are to be legitimate, neither one can be insistent. Simply stated, a person’s status as an end-in-itself is incompatible with the priority given to the impartial perspective presupposed by the claim that there is an *insistent pro tanto* reason to promote the good. Similarly, the rejection of *any* sort of *pro tanto* reason to promote the overall good is incompatible with the agent-neutral value that arises from our bargainers’ concern to preserve persons’ status as beings who are worthy of concern. (It is worth noting, too, that this very dimension of our bargainers’ motivations that leads to the acceptance of the non-insistent *pro tanto* reason is an indispensable component of a successful defense of contract-based moderate morality. For it is this belief that persons are worthy of concern that generated the threshold modifications to constraints, and thus enabled us in the first place to defend a non-absolutist – and thereby plausible – version of moderate morality.)

In conclusion, then, we see that the bargainers’ personhood-respecting motive leads to the acceptance of the validity of both agent-relative and agent-neutral values as sources of reasons for an agent’s actions. Thus, our moderate can agree with Kagan in accepting the *pro tanto* reason to promote the greater overall good (as a consequence of her commitment to persons’ worthiness of concern), while still defending options (as a necessary consequence of her belief that persons are ends-in-themselves). These options, too, are compatible with the bargainers’ insistence on the inviolability of persons and with their commitment to promoting self-interest (expressed in their appeal to cost as an argument in favor of options).

III. Concluding Remarks

If the contract-based account I have sketched in this thesis is plausible, then it seems that Shelly Kagan’s claim that the moderate cannot appeal to contractarianism as a means of defending moderate morality is mistaken. For I have shown that if (as may reasonably be assumed) a set of rational bargainers is motivated not only to arrive at a

³⁹ Kamm 1992, p. 355.

system of morality that maximized overall well-being, but also to formulate a set of rules that properly recognized and respected their moral status *qua* persons, then they would agree to a system of both genuinely deontological constraints that yielded principled exceptions when necessary *and* options. In short, my bargainers would adopt a genuinely *moderate* morality, casting serious doubt on Kagan's claim that any appeal to the contract approach will yield only extremist morality.

The reason this contractarian account succeeds in grounding a moderate morality while (in Kagan's mind, at least) all others fail is no doubt due to the addition of a second dimension to the bargainers' motivation – the personhood-respecting motive. The extent to which this motive is a plausible one for rational bargainers to adopt, then, is the extent to which we might hope for a contractarian defense of moderate morality after all. To see why the bargainers' personhood-respecting motivation provides such a promising means of founding a full-ledged moderate morality, simply consider, for example, the correspondence between each of three aspects of the moral status of persons and the three commitments which Kagan identifies as distinctive of the moderate's position: the belief that persons are worthy of aid leads to an acceptance of the *pro tanto* reason; the belief that persons are ends-in-themselves leads naturally to the commitment to options; and the belief that persons are inviolable lends itself to the erection of constraints.

Frances Kamm has recently written that “we cannot permissibly ‘bargain away’ our moral status not to be treated in certain ways to increase our life prospects or minimize rights violations. Our moral status is inalienable.”⁴⁰ I hope to have demonstrated that if such an insight is developed into a full-blown contractarian theory, we will have available to us the means of defending moderate morality against Kagan's challenge to it in *The Limits of Morality*.

⁴⁰ Kamm 2000, p. 218.

BIBLIOGRAPHY

Gauthier, David. *Morals By Agreement*. New York: Clarendon Press, Oxford. 1986.

Kagan, Shelly. *The Limits of Morality*. New York: Clarendon Press, Oxford. 1989.

_____ *Normative Ethics*. Boulder: Westview Press, 1998.

Kamm, Frances. "Non-Consequentialism" in *The Blackwell Companion to Ethical Theory*, Hugh Lafollete, ed. Oxford: Blackwell Publishers, 2000.

Kamm, Frances. "Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status." *Philosophy and Public Affairs* 21, no. 4 (1992). Princeton University Press.

Nagel, Thomas. "Personal Rights and Public Space." *Philosophy and Public Affairs*, 1995, pp. 83-107. Princeton University Press.

Nozick, Robert. *Anarchy, State and Utopia*. Basic Books, 1974.

Rawls, John. *A Theory of Justice*, revised edition. Cambridge: Belknap Press, 1999.

Scanlon, T.M. *What We Owe to Each Other*. Cambridge: Belknap Press, 1998.

Quinn, Warren S. "Actions, Intentions, and Consequences: The Doctrine of Double Effect." *Philosophy and Public Affairs*, 1989. Reprinted in *Ethics: Problems And Principles*, John Martin Fisher and Mark Ravizza, eds. Harcourt Brace Jovanovich, 1992.

Vallentyne, Peter, ed. *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*. New York: Cambridge University Press, 1991.

Vaughn Bryan Baltzly

Bryan Baltzly hails from a small and very pleasant town in northwest Ohio called Wauseon. He attended Hillsdale College, in a small and pleasant town in south-central Michigan, for four years after high school. After earning a bachelor of arts degree there in 1999, with majors in English and philosophy, he came to small, pleasant Blacksburg to pursue an M.A. in philosophy from Virginia Tech. Finding academia quite to his liking, he then sought and secured admission to the Ph.D. program at the University of Maryland-College Park, very near to Washington DC – which is neither small nor particularly pleasant. There Bryan will adjust to urban life and continue his graduate studies through the “Committee on Politics, Philosophy, and Public Policy.” He hopes to earn both a master’s in public policy and a doctorate in philosophy, and then wield his over-educated brain in both the academic and the political arenas.