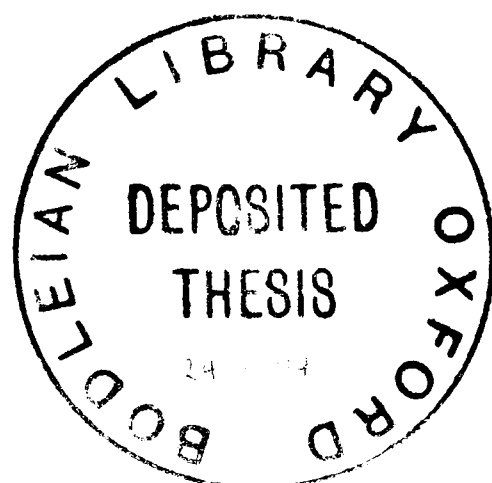


The Demands of Consequentialism

Timothy Paul Mulgan

**Thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy at the University of Oxford.**

Trinity Term 1994 [ie 1995]^{TT}



DECLARATION

I hereby declare:

- (i) that this thesis is entirely my own work;
- (ii) that this thesis contains no material which has ever been submitted for, or towards, any degree or other qualification;
- (iii) that previous versions of two sections of this thesis have been published, as follows: an earlier version of section 2.2.1 of Chapter TWO (pp. 47-56) appeared as "Rule Consequentialism and Famine" in *Analysis*,¹ while an earlier version of Chapter FOUR appeared as "Slote's Satisficing Consequentialism" in *Ratio*;²
- (iv) that this thesis contains approximately 63,300 words (including footnotes, but excluding the bibliography).



Timothy Paul MULGAN

1 February 1995

¹Mulgan 1994.

²Mulgan 1993a.

ACKNOWLEDGEMENTS

I would like to thank the following people for helpful comments on previous drafts of parts of this thesis: Tim Chappell, John Cottingham, Roger Crisp, Reid Hoffman, Brad Hooker, Rahul Kumar, Andrew Moore, Liam Murphy, Philip Pettit, Charles Pigden, Peter Smith, Nick Sparks, Mike Thrush, and an anonymous reviewer for *Ratio*.

Some of the material in this thesis was presented to audiences at the Wolfson Philosophy Society and the Reading University Philosophy Society. I am grateful to those audiences for their comments.

I am also grateful to Roger Crisp and Bernard Williams for patient and constructive supervision.

My greatest debt is to my principal supervisor, Derek Parfit, for his patient and insightful comments on innumerable drafts of all that follows, and for his support and encouragement.

ABSTRACT

The Demands of Consequentialism

Timothy Paul Mulgan

Wolfson College

DPhil

Trinity Term 1994

The thesis is an examination of the familiar objection that Consequentialism is unreasonably demanding (hereafter the Demandingness Objection). The focus is on attempts to construct a moral theory which avoids making unreasonable demands, without departing too much from traditional Consequentialism. The thesis is in two parts. In Part ONE, a wide range of contemporary forms of Consequentialism are examined, particularly the theories of Parfit, Brandt, Hooker, Murphy, Slote and Scheffler. It is argued that none of these is able to provide an adequate response to the Demandingness Objection. In Part TWO, a new Consequentialist theory is sketched. The core of this theory is a theoretical innovation: non-proportional accounts of the relationships between the values of outcomes, the costs faced by agents, and the rightness or wrongness of actions. It is argued that such accounts can provide the basis for a response to the Demandingness Objection. It is also argued that non-proportional elements can be incorporated into the theories of Scheffler and Parfit, and that the resulting theories are superior to the originals. It is concluded that the notion of non-proportionality is worthy of further exploration, and that the best possible Consequentialist moral theory is very likely to incorporate some non-proportional elements. It is also concluded that it would be premature to assume that Consequentialists will never be able to put together an adequate response to the Demandingness Objection.

TABLE OF CONTENTS

Declaration	ii
Acknowledgements	iii
Abstract	iv
Introduction	1
PART ONE: Past Solutions	
Chapter ONE: The Parfit Suggestion	6
Chapter TWO: Collective Consequentialism	34
Chapter THREE: Murphy's Cooperative Principle	87
Chapter FOUR: Slote's Satisficing Consequentialism	99
Chapter FIVE: Scheffler's Agent-Centred Prerogative	117
Part TWO: Towards a New Solution	
Chapter SIX: Non-Proportionality Introduced	147
Chapter SEVEN: a Non-Proportional Scheffler	164
Chapter EIGHT: Blameless Wrongdoing Revisited	181
Conclusion	201
Bibliography	203

INTRODUCTION

The thesis is an examination of the familiar objection that Consequentialism is unreasonably demanding. I call this the Demandingness Objection. Consequentialists claim that the right thing to do in any situation is the act with the best consequences. The Demandingness Objection is that this leaves the agent too little room (time, resources, energy) for her own projects or interests. My discussion will focus on the following story:

Affluent's Tale : Affluent is an affluent citizen of a developed country, who already makes significant donations to charity. She is sitting at her desk with her chequebook. In front of her are two pamphlets: one from a reputable international aid organisation, the other from her children's prospective music teacher. Affluent has enough money to either buy the music lessons or make a donation to the charity, but not both. Because of her love for her children, she buys the lessons, even though she knows that the money would have done far more good if sent to the charity.

The Demandingness Objection says that Consequentialism must condemn Affluent's behaviour, and that this is unreasonable.

0.1: Plan of the Thesis

There are three extreme responses to the Demandingness Objection. The first is to abandon Consequentialism altogether. The second is to accept that Consequentialism is very demanding, but to argue that we should just accept that morality is a very demanding business and make the required sacrifices. This is the response of Shelly Kagan.¹ The third response is to argue that, even as it stands, traditional Consequentialism is actually not

¹See Kagan 1989. A similar view is advanced in Singer 1972.

particularly demanding. For instance, Consequentialists have often suggested that, due to our limited knowledge of other peoples' affairs and other practical difficulties, it is unlikely that we would do more good donating our money to distant charities than spending it closer to home.²

I do not propose to spend much time on any of these three responses. As to the first, I accept that there are problems with Consequentialism. However, the intuitive appeal of the Consequentialist approach, combined with the fact that all current alternatives face problems of their own, suggests that it is still worthwhile to seek a response to the Demandingness Objection from within Consequentialism. As to the second response, I think it has little intuitive appeal. The idea that there are limits to the demands which morality may place upon us is a very compelling one. (Although, of course, there is room for considerable disagreement as to the precise location of those limits.) At the very least, it is certainly reasonable to seek a less demanding version of Consequentialism. As to the third response, it is fairly obvious that the best modern aid agencies will produce more good with whatever money we give than we could produce by spending that money within our own privileged circles. Our less demanding Consequentialism will not be found so easily.

We are thus searching for a theory which avoids making unreasonable demands, without departing too much from traditional Consequentialism. In Part ONE (Chapters ONE through FIVE), I evaluate a number of Consequentialist theories which have been prominent in the recent literature, and ask of each in turn whether it can provide an adequate response to the Demandingness Objection. My conclusions are mostly negative. No extant Consequentialist theory defeats our objection.

²This line of argument was discussed by Mill and Sidgwick, and has been defended more recently by Frank Jackson. (See Jackson 1991b and the references given there.)

In Part TWO (Chapters SIX through EIGHT), I sketch a new Consequentialist theory. I attempt to show that this new account provides a better response to the Demandingness Objection than any existing theory. The core of my proposed solution is a theoretical innovation: the idea of a non-proportional account of the relationship between the values of outcomes and the rightness or wrongness of actions. This idea is introduced in Chapter SIX. In Chapters SEVEN and EIGHT, I demonstrate that this device can also be combined with the theories of Scheffler and Parfit, which are discussed in Part ONE. I argue that the introduction of non-proportionality would improve either of these prominent theories.

0.2: Methodology

As far as possible, I seek to avoid any specific meta-ethical commitments. My methodology is similar to Sidgwick's "philosophical intuitionism". In each chapter of Part ONE, I begin by taking a pre-existing moral theory, which I then test against certain general intuitive standards which any acceptable moral theory must meet, as well as against our intuitive responses to particular cases. If the intuitions to which I appeal seem likely to be controversial, I usually attempt to justify them by appeal to less controversial assumptions. At some point, I reach assumptions which I do not defend. I simply try to state these as clearly as possible, and then trust that the reader shares them to a sufficient degree to find my argument persuasive. In Part TWO, I begin by developing a new theoretical device. I then seek to provide reasons to regard this device as at least worthy of serious consideration. I then test several theories which incorporate that device, using the methodology adopted in Part ONE.

0.3: Plan of Part ONE

There are many different forms of Consequentialism. We can classify these in terms of the following features:

(i) The Unit of Assessment: what does the theory assess? Possible units include acts, rules, motives, characters, lives.³

(ii) The Mode of Assessment: are the various units of assessment assessed directly or indirectly? In assessing acts, a Direct Consequentialist would look at the consequences of each act, whereas an Indirect Consequentialist might consider the overall consequences of the motive which produced the act. If the unit of assessment is assessed indirectly, then we must ask in terms of what it is assessed.⁴

(iii) The Comparison Class: phrases such as "the act with the best consequences" are ambiguous. The two extreme interpretations are as follows: "the act which, of those available to the agent, produces the best consequences, given that everyone else acts as they actually do", and "the act which is such that the consequences of everyone performing it would be better than the consequences of everyone performing any one of the other acts available to the agent". We can call these interpretations the Individual and Collective views respectively.

In Chapter ONE, I discuss an attempt (based on the work of Derek Parfit) to produce an acceptable Consequentialist theory without including either Indirect or Collective elements.

³For an explanation of the last of these, and for references to recent discussions of the other units of assessment, see Crisp 1992.

⁴For instance, acts can be assessed indirectly in terms of rules, or motives, or ...

In Chapters TWO and THREE, I discuss two recent forms of Collective Consequentialism (those of Brad Hooker and Liam Murphy respectively).

In Chapters FOUR and FIVE, I move on to theories which supplement a broadly Consequentialist theory with some distinctive non-Consequentialist element. I concentrate on Michael Slote's Satisficing Consequentialism (Chapter FOUR) and Samuel Scheffler's Hybrid View (Chapter FIVE).

PART ONE: PAST SOLUTIONS

Chapter ONE: The Parfit Suggestion

In this chapter I focus on the following claim:

The Parfit Suggestion : Consequentialists can reduce the force of the Demandingness Objection without resorting to either indirect or collective assessment (either of acts or of any other unit of assessment), by appealing to the notion of blameless wrongdoing.⁵

I shall discuss this suggestion in relation to Affluent's story. Parfit illustrates the notion of blameless wrongdoing by means of:⁶

Clare's Case : Most of the best possible sets of motives (judged in Direct Consequentialist terms) would include strong love for our children. Suppose that Clare has one of these sets of motives. Clare could either give her child some benefit, or give a much greater benefit to some unfortunate stranger. Because she loves her child, she benefits him rather than the stranger.

Consequentialists adopting the Parfit Suggestion will make three claims about this story:

(a) Clare performs the wrong act. Faced with a choice between two options, she fails to choose the one which would maximise the good;

⁵I call this the Parfit Suggestion because it was suggested to me by Derek Parfit, not because it is to be found explicitly in any of his writings.

⁶ See Parfit 1984, p. 32.

(b) Clare should not be blamed for her action. This is because she acts from a desirable set of motives. The only way Clare would have acted otherwise would be if she'd had different (less desirable) motives;⁷

(c) Clare is not to be blamed for her failure to alter her motives at some previous time. This is because, overall, things would have gone worse if Clare had taken such steps than if she hadn't.

Before going any further we should note a number of complexities and distinctions which often arise in discussions of Consequentialism. The first is the distinction between actual results and probable results. For instance, assume that A must press one of two buttons (X and Y). Independently, a random number between 1 and 100 is generated by a computer. If A presses button X and the number is 100, then an innocent person is electrocuted. If she presses X and the number is not 100, then no one is hurt. If A presses Y and the number is 100, then no one is hurt. On the other hand, if A presses Y and the number is not equal to 100, then an innocent person is electrocuted. Assume that A does press button X and the number selected is 100. A Consequentialism based on actual results would say that A had acted wrongly, as things would have turned out better if she'd pressed the other button instead. By contrast, a Consequentialism based on probable results would say that A had acted rightly, as pressing button X was far more likely to lead to good results than pressing button Y.

A second distinction is between objective and subjective probabilities. Returning to our previous example, assume that A is misinformed regarding the probabilities involved. She believes that if she presses button X there is a 99% chance of an innocent person being

⁷As we shall see, both throughout this chapter and in Chapter EIGHT, a great deal depends on how we gloss this last sentence.

electrocuted. She also believes that if she presses button Y there is only a 1% chance of someone being electrocuted. Assume that A presses button Y. A Consequentialism based on objective probabilities would say that she's acted wrongly, as in fact pressing button Y is 99% likely to lead to electrocution. On the other hand, a Consequentialism based on subjective probabilities would say that A has acted rightly, as she has performed the action with the lowest subjective probability of leading to harm.

Many objections to Consequentialism and many confusions within Consequentialism can be removed by careful attention to these and other related distinctions. For the purposes of this thesis, however, I wish to avoid these complexities as far as possible. Accordingly, unless otherwise stated, all references to the values of outcomes should be read as referring to their expected values, and all examples should be interpreted in such a way that objective and subjective probabilities coincide.

We need to distinguish three senses of the word "blame". A first distinction is between blaming as an act and blaming as a judgement. Acts of blame will include publicly reprimanding the agent, or otherwise finding fault with her. For the Consequentialist, an act of blame is treated exactly the same as any other act. If our assessment of acts is direct, then we should actively blame Affluent if and only if this would produce better consequences than anything else we might be doing at the time. Judgements of blame, by contrast, need not be treated like acts. We might (somewhat artificially) regard them as (consequenceless) assessments of an agent's actions, particularly where those actions are taken as indicative of her moral character. For the Consequentialist, it's easy to see how these two forms of blame might come apart, as it may be beneficial to blame someone whose acts and character are morally unobjectionable, whereas at other times we may judge that a person is worthy of blame, but decide that it would be for the best not to actively blame them.

Many Consequentialists elide the two notions of blame, and then provide an account only of acts of blame. For instance, Sidgwick offers the following utilitarian account of praise:

From a utilitarian point of view, as has been before said, we must mean by calling a quality 'deserving of praise' that it is expedient to praise it with a view to its future production. Accordingly in distributing our praise of human qualities on utilitarian principles, we have to consider primarily not the usefulness of the quality but the usefulness of the praise.⁸

Acts of blame are to be considered simply as acts and therefore, like any other acts, are to be assessed solely in terms of their consequences. I accept this as an account of acts of blame, but I doubt that it does full justice to the role which the concept of blame ought to play in our overall moral theory. The general appropriateness of blame in a particular situation does not seem to dissolve neatly into the consequences of a particular possible assignment of blame. For instance, it may often be expedient to blame a perfectly innocent person in order perhaps to set an example to others, but it would be odd to say that such a person was blameworthy. They don't deserve to be blamed any more than an innocent person killed in order to appease a hostile mob deserves to be killed. They're just in the wrong place in the wrong thought experiment. Furthermore, the effect of a particular act of blame will depend very largely on how it is phrased and on the precise relationship between the blamer and the blamee. Yet such factors are often irrelevant to the determination of blameworthiness.

I conclude that, if our Consequentialist theory is to adequately reflect the richness of our moral judgements, then it must leave some room for the concept of blameworthiness over

⁸Sidgwick 1907, p. 428.

and above the consequences of particular acts of blame. Of course, anyone who follows Sidgwick in rejecting this view is likely to find little merit in any response to the Demandingness Objection which places any weight on the distinction between wrongdoing and blameworthiness.

A second distinction is between two types of blame judgement, which I shall call judgements of oneself and judgements of others. We should ask, not only how we judge Affluent from the outside, but also how Affluent sees her situation from the inside: should she feel remorse, guilt or uneasiness? This first person point of view is a central feature of all our moral lives. We want a moral theory not just to help us make judgements of other people, but also to offer some guidance in our own deliberations. The question of what it is like to be Affluent is not remote from our experience. We are all Affluent.⁹ We all face dilemmas such as Affluent's on a regular basis. Many of these dilemmas take place in the calm reflective situations on which we have focused. A theory which offered overly-demanding advice in such situations would thus be seriously deficient.¹⁰

To simplify matters, I shall take Parfit to be claiming that Clare is not to be blamed in any of these three senses. We should not actively castigate her for her actions, we should not judge her to be morally deficient, and she herself should feel no guilt or remorse.

Parfit's position has been much discussed.¹¹ I don't want to enter into this debate here. For the sake of the present argument, I propose to accept that there are cases in which the best motives may produce a wrong action, and also to accept that Clare's story is such a case. I shall ask if Consequentialists can use these admissions to avoid demandingness.

⁹The scope of "We" being at least broad enough to include, inter alia, anyone who is likely to be reading this thesis.

¹⁰In Chapter EIGHT, I discuss the different senses of blame in more detail. I focus there on judgements rather than acts of blame.

¹¹See, for instance, Darwall 1986, Mendola 1986, Moser 1991, and Rabinowicz 1989.

1.1: Is Affluent Blameless?

If the Parfit Suggestion is to work, then we must show that Affluent's situation is analogous to Clare's. As we are assessing acts directly, Affluent's act is clearly wrong, as she knowingly and deliberately fails to maximise the good.¹² Affluent will claim that she is still blameless, as she acts in accordance with the best possible set of motives (or one of the best possible sets).

On the face of it, this seems rather implausible. It is relatively easy to construct other sets of motives which are such that, if Affluent had had them, things would have gone much better overall (and where this could have been predicted). For instance, if Affluent had completely impartial motives, she would donate large amounts to charity on a regular basis. Her own children (and probably herself and her friends as well) would no doubt suffer as a result, but this would be greatly outweighed by the good brought about her donations. It would thus appear that Affluent does not have the best possible motives.

Affluent and her supporters have three options at this point. First, they might argue that, although Affluent's case is not analogous to the first of Parfit's Clare cases, it is analogous to the second:¹³

Clare Two's Case : Clare Two could either save her child's life, or save the lives of several strangers. Because she loves her child, she saves him, and the strangers all die.

¹²Presumably, Affluent has taken account of the possible effect her choice may have on the future strength of her desirable parental motives. Ex hypothesi, any negative effect the charitable donation might have in this regard is more than outweighed by the benefit the donation would bring. On balance, the donation is the better act.

¹³ *ibid.*, p33.

Parfit concedes that, if at some time in the past Clare Two had given herself more impartial motives, then things would have gone better overall. However, when deciding which motives to have, she didn't know that her present situation would arise. The probability that Clare Two would end up in this type of life or death situation was sufficiently low that the expected value of the inculcation of parental motives exceeded the expected value of the inculcation of impartial motives. This is enough to absolve Clare Two from blame.

Fortunately, we don't need to assess Clare Two's defence. This is because her case is not analogous to Affluent's. Unlike Clare Two, Affluent cannot plausibly claim that she didn't know that her present situation would arise. The letter from the charitable organisation is hardly a bolt from the blue. Affluent has known for some time that she would face such choices. She must have known that, given her current motives, she would act very wrongly. She could have taken steps to change her motives. She didn't. She can thus be blamed for acting on her present motives, as these are not the best possible set.

Affluent's first reply fails. She cannot admit that her current motives are not the best, and then attempt to excuse herself. She must claim that she has the best possible motives. Her two remaining options represent two different forms this claim might take. The first of these is to amend our story as follows: Affluent is not an inhabitant of the world as it is now. Rather, she lives in an improved world of the (possible) future, where there is far less poverty, and people are much more generous. In that society, the best set of motives for Affluent are her actual parental ones.

As it stands, this reply is inadequate. We are interested in Affluent precisely because her situation is one which commonly arises in our own world. To escape the Demandingness Objection, Consequentialists must deal with a real-world Affluent, not one who lives in some idealised never-never land. Affluent might reply that the best motives for her to have

in our actual world are those which would produce the best consequences in that ideal society. In other words, Affluent could replace her individual assessment of acts with a collective one. Unfortunately, such a move constitutes an abandonment of the Parfit Suggestion, which aims to avoid collective assessments. Whatever the merits of Collective Consequentialism, it is not an option for Affluent at this point.¹⁴

Affluent's second option is to claim that she couldn't have given herself more impartial motives without the benefits of those impartial motives being outweighed by the negative effects of the inculcation process itself. (For instance, perhaps a more impartial Affluent would be a listless and morose figure who wasn't much good to anyone.) To test this new claim, we must construct a set of motives containing the highest degree of impartiality which Affluent could have successfully inculcated in herself without any negative side-effects. Affluent would then have to demonstrate that a theory which required her to act in line with that set of motives would not be not unduly demanding.

The question of what is psychologically possible is rather complicated, and I do not hope to adequately address it here. However, Affluent's claim does seem dubious. In general, people do seem capable of having very demanding impartial motives. Although Affluent is already much more impartial than most, as she gives a significant amount to charity, it seems likely that she could have trained herself to give more, if she had chosen to do so.

¹⁴I discuss the merits and demerits of Collective Consequentialism at length in Chapters TWO and THREE. Collective assessments are likely to be less demanding than individualistic ones, but they face other objections which the latter avoid. In addition to general worries regarding Collective Consequentialism, we would also have to ask whether a direct collective assessment of motives is compatible with a direct individualistic assessment of acts. This question has both a practical and a theoretical dimension (though these overlap). The practical problem is that, although we want some divergence between the best act and the best motive - otherwise we would not have been attracted to the Parfit Suggestion in the first place - this present combination may produce too much divergence. Blameless wrongdoing may become too common. The theoretical problem is that there may be a tension between our rationale for a collective assessment of motives, and our rationale for assessing acts individually. It may be difficult to justify assessing the two features in such radically different ways. This would lead us to another issue, the possibility of combining our direct collective assessment of motives with a direct collective assessment of acts. This would take us still further from the original intention of the Parfit Suggestion.

Even if we grant Affluent's claim, her approach leaves us too much at the mercy of empirical facts. We certainly want the demands of morality to be sensitive to the limits of psychology. Morality should not demand what is psychologically impossible. However, should it demand all that is psychologically possible? Perhaps those who are capable of greater impartiality should be more impartial than others, but are they required to be as impartial as they possibly can?¹⁵ This seems too demanding. An adequate response to the Demandingness Objection should leave us with some breathing space between the demands of morality and the limits of psychology. We shouldn't always be blamed or blame ourselves for failing to be the best we possibly could.

Furthermore, even if Affluent acts on the best set of motives, she might still be blameworthy. At some point in the past, Affluent decided to have children.¹⁶ She knew that this decision would result in her being placed (almost continuously over a long period of time) in situations in which the best possible motives she would be able to develop would result in her acting wrongly. Alternatively, things would have been much better overall if Affluent had devoted all the time, money and energy she's lavished on her children to alleviating the sufferings of strangers. Affluent can be blamed for her decision to have children because she failed to make the world a much better place than it would otherwise have been.¹⁷

¹⁵ Subject only to the proviso, noted above, that the benefits of more impartial motives must be weighed against any costs necessitated by their inculcation.

¹⁶ Of course, Affluent might have become pregnant against her will and had no access to termination facilities, but we shall assume that this is not the case.

¹⁷ We should note that the question of whether Affluent's decision to have children constitutes a failure to make the world a much better place than it might have been may depend upon how we resolve a controversial matter in the Utilitarian theory of value. Derek Parfit presents two rival interpretations of the Utilitarian principle that we should seek the greatest happiness of the greatest number (see Parfit 1984, pp. 386-387):

The Impersonal Average Principle : If other things are equal, the best outcome is the one in which people's lives go on average best.

At this point, we need to distinguish two claims which proponents of the Parfit Suggestion might make:

(1) Affluent was right not to take steps in the past to give herself different motives;

(2) Affluent is right not to take steps now to give herself different motives.

These two claims are distinct. In particular, (2) seems more plausible than (1). Affluent might claim that she has the best motives for her present situation, but she can't deny that she should have given herself different motives which would have landed her in a different situation. If she could establish (2) without (1), then we might say that Affluent was partially blameless. There would probably be no point in our castigating her, as we don't

The Impersonal Total Principle : If other things are equal, the best outcome is the one in which there would be the greatest quantity of whatever makes life worth living.

These two interpretations come apart in cases where different numbers of people would exist in different possible outcomes. The decision whether or not to have children is clearly a case of this kind. Under the Average Principle, Affluent's decision to have children is almost certain to count as a failure, as by devoting her life to improving the lives of people who would have existed anyway she would have raised the average level of well-being far more than she does by having a few very well-off children of her own. On the Total Principle things are more complicated. Simply by creating additional lives which are worth living Affluent increases total value. On this view, making extra happy people is, other things being equal, always good. However, even in terms of the Total Principle, Affluent's actions come at considerable cost, as she could have made a significant contribution to the well-being levels of others. It is quite possible that, on balance, the total level of whatever makes life worth living would have been higher if Affluent hadn't had children, as the increase in Third World happiness would have outweighed the loss of her children's well-being. This would only fail to be true if the very existence of each individual life counted for a very great amount of value, irrespective of the quality of that life. On this view, anything Affluent does to increase the global population will be desirable. However, such an extreme view strikes me as absurd and I am unable to see any reason to adopt it.

The situation is further complicated by the fact that the effects of Affluent's donations would depend on the policies of Oxfam and other relief agencies. Yet these agencies themselves must choose between the Average Principle and the Total Principle, especially in their attitudes to family planning. In particular, if relief agencies adopt the Average Principle, then under the Total Principle it may be best for Affluent to donate nothing at all, as any donation might be used by the agency to improve overall average utility at the expense of total utility. On balance, it seems likely that Affluent would have maximised total value by not having children, so we can plausibly proceed without choosing between the Average Principle and the Total Principle. On either principle Affluent ought not to have had children.

want her to change her motives now, but we would still judge that her present lack of impartiality was blameworthy.

However, this partial blamelessness would not be enough. Even if Consequentialism didn't make great demands in Affluent's present situation, it would still be very demanding in what we may call the pre-Affluent situation, as no affluent parent would ever be allowed to have children. Most of us would regard this demand as excessive.

Affluent has no compelling reply to the charge that her motives are not the best. We cannot use the notion of blameless wrongdoing to provide a complete solution to the Demandingness Objection. However, this does not mean that the notion has no role to play. Perhaps the best response to the objection will be a combination of several different replies each of which serves to weaken the force of the objection, though none is a complete answer on its own. Blameless wrongdoing might be one part of that composite response.¹⁸ It is therefore worth asking whether blameless wrongdoing would be of any use in cases where it could be established.

1.2: The Nature of Blameless Wrongdoing

In this section, I explore the nature of blameless wrongdoing by discussing four objections to Parfit's notion: that it produces absurd results, that it is irrational, that it undervalues acts, and that it is phenomenologically implausible.

¹⁸Indeed, in Chapter EIGHT, I make quite extensive use of a version of Parfit's distinction between wrongness and blameworthiness.

1.2.1: An Absurd Result

Consider the following story:¹⁹

The Warlord's Tale : Warlord wishes to rule the world. She begins by starting an extremely bloody civil war in her own country. She then invades a neighbouring country. The rest of the world is so alarmed by Warlord's ambitions that a new spirit of international cooperation is forged to enable the United Nations to deal with the threat. Once forged, this new spirit persists. As a result, several potentially far more successful megalomaniacs are nipped in the bud. In fact, the world is a better place over all than it would have been if Warlord hadn't embarked on her career of barbarity.

This story is not entirely implausible.²⁰ Yet it creates havoc for this use of Parfit's account. Warlord acts from a desire to rule the world. If Warlord had altered her motives at some earlier date, then the world would have been a worse place overall. It would have been wrong for Warlord to alter her motives. Therefore, her behaviour is praiseworthy. Furthermore, if Warlord had altered her motives so as to allow her to remain peacefully at home on her farm instead of embarking on a campaign of mayhem and slaughter, she would have been blameworthy, as the world would have been a worse place than if she'd lived the latter life. These judgements seem pretty absurd.

The most obvious response to this objection would be to distinguish between actual outcomes, objectively probable outcomes and subjectively probable outcomes. If we accept

¹⁹I owe this objection to Derek Parfit.

²⁰It is interesting to note that an extreme version of precisely this form of argument is to be found in some theistic responses to the problem of evil, where it is argued that the world would be a worse place without its current (and past) stock of ruthless dictators. This amounts to the claim that we can tell a version of Warlord's Tale in the case of every particular atrocity by every individual dictator. This general claim certainly is implausible. However, our use of Warlord's Tale requires only the much weaker claim that such a tale might be true of one particular series of atrocities performed by one individual dictator.

that only the last are relevant to questions of blame, then we will say that an agent is blameworthy for acting on a particular set of motives if and only if the subjective expected value of the history of the world in which she has those motives is less than the subjective expected value of the history of any world in which she has some alternative set of motives which had been available to her at some previous time. Presumably, Warlord does not expect her campaign to result in the strengthening of the United Nations. The desirable consequences of her warlike motives are entirely unexpected. Therefore, her actions are blameworthy, as she could have expected to create uncompensated mayhem.²¹

To get a pure case of our current objection, we must focus on a story in which Warlord does expect her campaign to end in defeat. We need to distinguish three versions of this new tale:

Warlord's Glorious Tale : Warlord expects to be defeated by the United Nations, but retains her warlike motives regardless. This is because she believes that it is better to die gloriously in battle than to live under the yoke of Western cultural imperialism.

Warlord's Selfish Tale : Warlord expects her regime to be overthrown. However, she does not expect this to occur until after her own death. She retains her warlike motives as this will make her own life go best.

Warlord's Diplomatic Tale : Initially Warlord has peaceful motives. However, she is concerned by the present weakness of the United Nations. Therefore, she acquires warlike

²¹This line of argument raises numerous problems surrounding the distinction between actual outcomes, objectively probable outcomes and subjectively probable outcomes. I discussed these at length in my honours dissertation (Mulgan 1990), and I shall side-step them here. It is worth noting that the problem of unexpected desirable consequences arises just as strongly in the context of isolated assessments of acts. It is thus not a specific problem for the assessment of motives per se.

motives and establishes her dictatorship in order to unite the international community against her.²²

Each of these tales is more complicated than our original. However, it does seem clear that, in both the Glorious Tale and the Selfish Tale, Warlord is to blame for her actions. In the Diplomatic Tale, by contrast, it is less clear that Warlord's actions are blameworthy. Certainly, many non-Consequentialists will be horrified by her behaviour, because it reflects exactly the sort of Machiavellian Consequentialist reasoning which such people find most objectionable. However, for precisely that reason, Consequentialists are likely to be fairly sympathetic to Warlord's behaviour. Once we reject the distinction between doing and allowing, we see in Warlord's actions a courageous self-sacrifice of the highest order.

As my project is to find a Consequentialist response to the Demandingness Objection, I shall proceed on the assumption that we are at least reasonably sympathetic to Warlord's behaviour in the Diplomatic Tale. Let us focus, then, on the contrast between the Diplomatic Tale and the Selfish Tale. Our intuitive reaction to the two tales is markedly different. However, in both tales Warlord has the same motives and performs the same actions with the same consequences. So long as we assess motives in terms of the consequences of the agent's possessing that motive, the problem of unsavoury motives with desirable consequences will remain. We need to construct an assessment which is more sensitive to the structure of the motive.

One such assessment naturally suggests itself, under which a motive is desirable only if it was chosen for a good reason. For instance, if, as in the Diplomatic Tale, Warlord chose

²²It is tempting to reply at this point that Warlord could achieve her goals just as easily by only pretending to have warlike motives rather than by actually causing herself to have such motives. However, this need not be the case. Perhaps Warlord will only be truly successful and convincing in her new role if she really has genuine warlike motivations. Furthermore, if it made no difference at all whether or not Warlord had warlike motives, then, on Parfit's account, such motives would be no more or less blameworthy than peaceful motives. This would still be a fairly odd conclusion.

her warlike motives because these would produce the greatest possible good, then she is to be praised for acting on those motives. By contrast, if Warlord chose her warlike motives on selfish grounds, as in the Selfish Tale, then it may be appropriate to blame her for acting on those motives.

On this account, we are primarily interested in the past choice between motives. If this choice was made for the right reasons, then the motive itself is desirable; if not, not. We still lack an account of what those right reasons are. Perhaps, at this level, we can simply say that the right reason for choosing a motive is that the motive will produce more good than any other motive available to the agent at that time.

The obvious alternative for Parfit would be to say that a desirable motive is one which is chosen as a result of a prior motive which is itself desirable under the original Parfitian criterion. Warlord's choice of motives in the diplomatic tale is motivated by her desire to make the world safe from dictators. If she hadn't had that motive, things would have gone worse, as she wouldn't have caused herself to have warlike motives and therefore wouldn't have embarked on her campaign. Therefore, it would have been wrong for Warlord to cause herself to lose her prior motive. Her present actions are not blameworthy.

There are a number of problems with this response. The first is that it threatens an infinite regress. We could, after all, construct a story in which an agent's prior choice of motives proceeded from a disreputable motive, but in which things would have gone worse if she'd lacked that disreputable motive. Indeed, our Selfish Tale seems to be just such a case.

Warlord's retention of warlike motives is motivated by her selfish desires. If she had lacked such desires, she might have opted to remove her warlike motives, but then things would have gone worse overall. Therefore, under the original Parfitian criterion, Warlord

would have been wrong to alter her selfish prior motives. Her present actions thus cannot be blameworthy.

Of course, we could then require a third order motive which met the Parfitian criterion, but then we could always construct another story in which some third order motive was present for disreputable reasons but nonetheless produced desirable effects. At each stage the Parfitian account looks less and less credible.²³

A second problem with this response is that it seems ad hoc. If we are rejecting the Parfitian criterion at the level of present motives, why apply it to the motives behind the prior choice of motives? Why not just reject the criterion altogether?

We are left then with an account on which a desirable motive is one which was chosen for good reason, where the latter notion is not itself explained in terms of the Parfitian criterion. Even this more general position, however, has its problems. The first is as follows. What do we do if no explicit better choice was made? What if the agent never explicitly chose to have or retain the motives under which she now acts?²⁴

A second problem for this revised account is that it seems redundant. Instead of applying our new account of good reasons to the choice between motives, why not apply it directly

²³Derek Parfit has suggested to me that we may be able to stop this regress if we reach a point in the past at which the best motive for Warlord to have would have been an Act Consequentialist one. We can then say that Warlord's present motive is acceptable if and only if it's one which Warlord would have had now if she had acted on Act Consequentialist motives at that time. However, it's not obvious that such a terminal point will be reached in every practical case. There may be no point at which things would have gone best if Warlord had had Act Consequentialist motives.

²⁴We should note that this question is much more pressing here than the analogous question for Parfit's original view. The original theory does ask what would have happened if the agent had chosen to alter her motives. However, because it assesses different motives solely in terms of the consequences of the agent's possessing those motives, Parfit's original view is much more likely to be able to give definite answers in cases where no prior choice of motives occurred. But it will usually be much easier to tell which set of motives would have led to better consequences than to tell what motivations the agent would have had if she had chosen between those motives. This point is related to the greater indeterminacy of backtracking counterfactuals than standard counterfactuals, an issue which is discussed at length in Chapter EIGHT.

to the choice between acts? Why not say directly that a good act is one chosen for a good reason? If a good reason is one which produces good consequences, then we're back to the original Parfitian account. If, on the other hand, the good reason for action is the desire to promote the good, then our distinction between wrongdoing and blameworthiness is in danger of collapsing into the distinction between subjective and objective wrongness, as a wrong act will be one which doesn't maximise the good, whereas a blameworthy act will be one which the agent performed without the intention to promote the good. Neither of these alternatives seems desirable.

A third, and related, problem for this more complicated account is that it loses the elegant simplicity of Parfit's original. In the original account, an act is blameworthy if, and only if, it would have been wrong for the agent to lose the motive from which it is performed. Blameworthiness can thus be defined in terms of wrongdoing. This is no longer possible in the revised account, as the grounds for assessing a good reason for choosing between motives are completely different from the grounds for assessing a choice between actions. The link between wrongdoing and blameworthiness has thus been broken. This loss injects a sense of arbitrariness into the overall account and significantly reduces its appeal.

I conclude that this revised Parfitian account of blameworthiness is not adequate. Defenders of blameless wrongdoing appear to have no acceptable response to the problems raised by Warlord's tale.

1.2.2: Blameless Wrongdoing is Irrational

On what I shall call the Standard View of Rationality, it is rational for an agent to do what she has most reason to do. It may seem that, on this view, Affluent behaves irrationally, as her deliberations don't seem to provide any reason to choose the music lessons. The fact

that one has a motive to do x is not the sort of reason for doing x which one could present to oneself in a moral deliberation.

This objection is too swift. The Standard View is a very thin theory of rationality. It gives us no guidance as to what can or cannot count as a reason. To construct a full theory of rationality, we must supplement the Standard View with an account of what an agent has most reason to do. Some candidates are the following:

The Present Aim Account : an agent has most reason to do whatever will best achieve her present aims, whatever those may be;

The Critical Present Aim Account : an agent has most reason to do whatever will best achieve her "ideal present aims", where these consist of: (i) certain aims which are deemed to be "rationally required" (whether or not the agent has them), plus (ii) the agent's actual present aims, minus (iii) any of the aims in (ii) which are deemed to be irrational;

The Self Interest Account : an agent has most reason to do whatever will make her life go best overall, whatever her present aims may be;

The Dominant Morality Account : an agent has most reason to do whatever is morally right, whatever her aims or interests may be.

I don't propose to decide between these accounts here. I wish merely to note that different accounts give different assessments of Affluent. On the Present Aim Account, Affluent is not irrational, as she does what will best promote her aim of educating her children. As, ex hypothesi, this is Affluent's strongest aim at the time of her decision, she chooses rationally.

The Critical Present Aim Account is less straight-forward. If we hold that it is irrational to educate one's children at the expense of the common good, then we will regard Affluent's choice as irrational. Otherwise we will get the same result as for the Present Aim Account. However, if we are to admit any aims at all other than those of pure self-interest and of morality, then the desire to educate one's children seems a very strong candidate. The Critical Account must admit some such aims, otherwise it will collapse into either the Self Interest Account or the Dominant Morality Account. So the Critical Present Aim Account is almost certain to hold that Affluent acts rationally.

On the Self Interest Account, Affluent may be in trouble. Perhaps, by spending the money on her children rather than on herself, she acts contrary to her own long-term self-interest. She would thus be acting irrationally. However, we should note that donating the money to charity would be even more irrational, as the contribution to Affluent's self-interest would be far less.

Only a Consequentialist who adopts the Dominant Morality Account must clearly condemn Affluent's choice as irrational, as she doesn't do what is right. If Consequentialists who adopt the Parfit Suggestion want to say that it is at least as rational for Affluent to buy music lessons as for her to donate to charity, then they are free to also adopt the Standard View of rationality so long as they don't combine it with the Dominant Morality Account of what an agent has most reason to do. While the other accounts may face problems, none of these affect the Parfit Suggestion in particular.²⁵

²⁵It is worth noting, in passing, that advocates of the Dominant Morality Account need not be dismayed either. While, on their view, Affluent is guilty of irrationality, her's is a case of what Parfit calls "rational irrationality", which arises where an agent causes herself to be irrational for some rationally defensible reason (see Parfit 1984, pp. 12-13). Rational irrationality is analogous to blameless wrongdoing. Indeed, on the Dominant Morality Account, the two will be the same thing, as rationality and morality have been equated. It would therefore be misleading to regard the charge of irrationality as a separate objection to Affluent's position.

1.2.3: Parfit Undervalues Acts

It might be objected that, by placing so much weight on the notion of blameworthiness, Parfit underestimates the significance of the direct moral assessment of acts. To see the force of such an objection, we might consider an extreme existentialist account of moral rationality, under which human decisions are taken to be totally free. At the moment of choice, nothing outside the choice itself determines what I will do. My past, environment, character, and motives are all insufficient to make me do one thing rather than another. I am always free to identify myself with a particular motive or not, to perform a particular action or not. To avoid responsibility for my present choice by blaming it on some "external" factor such as my own character or motives is to practice "bad faith". Existentialists would say that, when she says she's not to blame, Affluent indulges in bad faith. Her act does not result just from her character. It results from her having chosen, at the moment of choice, to acknowledge that character and identify herself with it. She could have done the right thing, and should be blamed for her failure to do so.

There is much to object to in this extreme position.²⁶ However, the existentialist picture does have some appeal as a phenomenological account of what it is like to make a calm reflective choice. We certainly do feel free to act otherwise than we do. Affluent, sitting in her study, will feel that she could write a cheque to charity, despite the strength of her parental motivations.

Affluent might claim that she's been misunderstood. She never said that she couldn't make the charitable donation. She could have done so, because she would have done so if she'd

²⁶It may also appear that the existentialist position is incompatible with any form of Consequentialism. However, I believe that an existentialist account of freedom must be supplemented by some positive moral theory as to what we should do with that freedom. I see no reason why Consequentialism could not play that positive role.

wanted to. But she didn't want to. Furthermore, it's right that she didn't want to, as things would be worse overall if she had. That's why she's not to blame and feels no remorse for her actions, even though she admits that she's done the wrong thing.

I doubt that existentialists would be satisfied with this. They would probably reply that, if Affluent genuinely believed both that she could have made a donation to charity and that she should have done so, then she would feel remorse at her omission. Her failure to feel remorse thus suggests that she doesn't really believe what she claims to believe. Affluent is being insincere, refusing to take full responsibility for her actions. Existentialists might regard this refusal as another form of bad faith. They would continue to insist that Affluent can only avoid blame by demonstrating that her act is not wrong.

Existentialists base their objection to Affluent's claim on a radical account of human freedom, which we probably do not share. However, anyone who believes that acts are a central unit of assessment will sympathise with this existentialist unease. They may feel that any agent who believes that she's done the wrong thing would feel at least some remorse. Affluent's claim that, while she sincerely believes that she's acted wrongly, she doesn't feel any remorse, may seem incredible.

Affluent will reply that, in general, wrongdoers should feel remorse. She herself would ordinarily feel guilty if she'd knowingly done wrong. However, her present act is no ordinary case of wrongdoing. It is an exceptional case, to which the presumption of remorse does not apply. What makes it exceptional is precisely the fact that Affluent acts from the best possible motives.

We may well accept that desirable motives lessen the degree of remorse required of a knowing wrongdoer. It is less clear that we would agree that such motives justify a total

absence of remorse. We may still believe that, even with the best motives, a wrongdoer should feel some unease. The crucial question will then be: is the degree of remorse we expect from a knowing wrongdoer (with the best motives) greater than the degree of remorse (if any) we feel it would be reasonable to demand from Affluent? If we answer this question in the affirmative, then we will only be able to defeat the Demandingness Objection by demonstrating that Affluent's act is not wrong.

At this stage we may be able to do little more than trade intuitions. Proponents of the Parfit Suggestion may claim that they have bridged the gap between the demands of Consequentialism, on the one hand, and what we feel is appropriate in Affluent's case on the other. I am inclined to feel that a significant gap remains. However, the notion of blameless wrongdoing certainly appears to have narrowed the gap, which is enough to weaken the force of the Demandingness Objection.

1.2.4: An Implausible Phenomenology

Perhaps the most striking fact about Affluent's situation is the room it leaves for reflection. Affluent has time to take a step back and examine the deliberative structure of her own moral decision making. Affluent can assess her motives and values, not just apply them. We can see this even more clearly if we revise our story as follows: Affluent sits at her desk one Sunday evening to organise her tasks for the coming week. She doesn't have to write her cheque until Friday. This leaves ample room for soul-searching.²⁷

We must remember that (for the purposes of this chapter) Affluent is not an Indirect Consequentialist: she doesn't define the right act as that which follows from the best

²⁷To further complicate matters, we might assume that Affluent is a card-carrying Consequentialist, familiar with the literature regarding the Demandingness Objection.

motive. She explicitly acknowledges that the right act is the act with the best consequences. Yet that act is the donation to charity. Instead of choosing the best act, Affluent chooses the act which is in line with the best motive. It's tempting to say that she chooses the best motive, and then acts in accordance with it. However, this would be misleading. Motives are not things which can be chosen immediately at will.²⁸ Motives must be acquired, cultivated, inbred, etc. When she comes to write her cheque, Affluent either has parental motives or she does not. Her decision is not whether or not to have the motive, but whether or not to act upon it.

We can perhaps best discuss this feature of Affluent's story in relation to a class of problem-cases which I shall call cases of "self-moulding" (or "moulding" for short). Moulding comes in many different shades. I shall focus on two of these, which are:

Inculcating : at time t , an agent (A) affects her decision at time $t+1$ by altering the motives and desires upon which that decision is based;

Training : at time t , an agent (A) acts to "train" her future self to do x . This act affects her decision at time $t+1$ by altering the procedure by which that decision is reached.

These two strategies will often go together. Sometimes they will be indistinguishable. However, they may also conflict. This is because, as a general rule, motives and desires produced by inculcation will be less responsive to deliberative reflection than decision-procedures instilled by training. In particular, problems may arise when an optimal moulding strategy produces sub-optimal choices in particular situations.

²⁸If they were, much of the point of the distinction between acts and motives, especially as it is used by Consequentialists, would be lost.

Some situations of this type will be more problematic than others. In particular, different moulding strategies will produce different degrees of awareness. At one extreme, we have cases of spontaneity. Some anti-Consequentialists have stressed the importance of acting "spontaneously" in certain situations. Consequentialists have replied that Consequentialism can accommodate spontaneity, by allowing people to train themselves to act spontaneously in the right sorts of situations. Now what makes a decision a spontaneous one, is precisely that it is not produced by conscious deliberation. If A is thinking "Now, this is one of those situations where spontaneity is desirable. What would a spontaneous person do here? Oh yes, she'd do x. I must do x", then A is clearly not behaving spontaneously.

This absence of reflection shows that the phenomenology of such decisions is relatively unproblematic. Seeing in advance that, at time t, one will choose a sub-optimal option raises no special difficulties, as, ex hypothesi, any other available strategy would produce even less desirable results overall. Similarly, reflecting on a sub-optimal decision in one's past, while not the most pleasant of activities, is not theoretically problematic. Sub-optimal decisions resulting from moulding are only peculiarly problematic at the time they're made. The problematic thought is not: "My life as a whole contains moments of sub-optimality". It must do this. Rather it is: "I am now deciding to take a sub-optimal option". This thought can only arise when the agent is reflectively aware of her decision.²⁹ For this to be the case, the following conditions must be met:

²⁹I wish, at this point, to parenthetically make a small strategic point. Discussions of Consequentialist moulding often centre around two familiar objections to Consequentialism:

The Calculating Objection : no moral agent would calculate in this situation. Consequentialism requires calculation in all situations. Therefore, Consequentialism is inadequate;

The Implausible Phenomenology Objection : Consequentialist moulding must sometimes require the agent to choose a sub-optimal option, and yet be aware that she is doing this. Such choices are phenomenologically implausible. That is, no agent could choose in the required fashion and still have a coherent awareness of her own deliberations. Therefore, Consequentialism cannot provide a plausible moral theory.

Both objections should be taken seriously. However, in any particular situation, at least one of them must be redundant. Take any moral decision, D (made by A at time t). If the Calculating Objection is valid

(1) The moulding strategy leads to sub-optimal decisions of which the agent is simultaneously reflectively aware.

(2) Some of the decisions in (1) are ones where the agent is aware that her decision is sub-optimal.

(3) Some of the decisions in (2) are ones where the agent is aware that her decision results from a Consequentialist strategy of moulding.

(2) will be especially problematic if the agent knows which option is optimal, and then chooses another one. In other words, the thought "I am now not choosing to x, even though x-ing would be the optimal thing to do" must, in some sense, be available to A. Such an agent is, by definition, not a Consequentialist agent.³⁰

Condition (3) may be even more problematic than Condition (2), as the agent chooses in an anti-Consequentialist manner, despite being aware that she only does so because of a decision of her previous (Consequentialist) self. This situation, if it occurs, will be a confusing one to be in. Yet, for Consequentialists who follow the Parfit Suggestion, this will be a common experience. For instance, Affluent is aware that she is making a sub-optimal choice. She also knows exactly what the best option would be. So Affluent satisfies both Conditions (1) and (2). Furthermore, Affluent's plea that she not be blamed for her failure to alter her motives, suggests that she acquired her present motives as a

against D, then the Consequentialist can only dissolve it by providing A with a strategy which ensures that, at t, A decides without calculating. However, I have attempted to show that, if deliberation is avoidable, then so is contemporaneous awareness. So A cannot be aware, at t, that she is making a sub-optimal decision. Therefore, the Implausible Phenomenology Objection cannot arise. In any given situation, then, anti-Consequentialists must choose one objection or the other. They cannot have both.

³⁰This may look like a fatal problem for Consequentialism. However, it is not. The fact that, at time t, A is not a Consequentialist does not show that she was not a genuine Consequentialist when she embarked on the strategy of moulding which lead her to be as she is at t.

result of some strategy of moulding in her past.³¹ So she will satisfy Condition (3) as well.

The presence of moulding underlines the significance of our third sense of blame: remorse, guilt and other feelings of self-blame. Our reply to the Demandingness Objection must do justice to our sense of which feelings are appropriate to a situation like Affluent's. We may discover that certain feelings of remorse and unease, which we would expect in cases of sub-optimality produced by moulding, would not be appropriate for Affluent. We could not then regard Affluent as a purely Consequentialist self-moulder. If such moulding is a necessary accompaniment of blameless wrongdoing, then we will be unable to explain our response to Affluent simply in terms of blameless wrongdoing.³²

1.3: Conclusions

Our discussion has been too sketchy to allow definite conclusions to be drawn. We began with the Consequentialist claim that Affluent does the wrong thing. The Demandingness Objection rejected this claim as unreasonable. The Parfit Suggestion seeks to retain the original claim, by arguing that, once we've seen that Affluent should not be actively blamed, that we shouldn't think ill of her character and that she herself should feel no guilt or remorse, we won't regard it as unreasonable to say that she does the wrong thing. Against this, I argued that Consequentialists will be unable to avoid blaming Affluent, as she probably doesn't have the best possible motives. I also argued that cases such as the

³¹We need not assume that Affluent was once a Consequentialist who deliberately became a non-Consequentialist. She may never have been fully Consequentialist. However, at some time in the past when she could have made herself more Consequentialist she decided not to do so, for Consequentialist reasons. This would be a case of what we might call "Negative Moulding", where the agent influences her future motives by deciding not to embark on a positive moulding strategy.

³²Moulding also brings out the possibility that questions of the continuity of personal identity over time may become relevant to our assessment of Affluent's tale. Should Affluent be worried that she's being manipulated by her previous self? This worry may be particularly acute for someone, like Parfit, who has a weak notion of personal identity (for a discussion of which, see Parfit 1984, pp. 199-347). If Affluent is very different from her previous self, she may feel she's being manipulated by someone else.

Warlord's Tale raise more general doubts regarding this use of Parfit's distinction. Finally, I argued that even if Affluent were blameless, this might not be enough. Those who give priority to the assessment of acts will not be satisfied with anything less than the admission that Affluent does no wrong. Others, while admitting the significance of other units of assessment, may argue that a complete solution to the Demandingness Objection cannot be built on so stark a conflict between different moral assessments. Affluent situations are common enough, and the degree of awareness they involve is acute. We may be suspicious of a moral theory which tells us that, in all these cases, a morally blameless person is knowingly doing wrong.³³

Of course, suspicions are not arguments. Perhaps the Parfit Suggestion can be made to work; or perhaps it will provide part of a wider solution to the Demandingness Objection. However, sufficient doubts have been raised about the suggestion to justify the search for

³³Similar considerations also provide an argument against another familiar Consequentialist approach, which we do not have time to discuss in any detail: namely, the two level theory of R. M. Hare. (See, for instance, Hare 1981.) Hare distinguishes two levels of moral thinking: critical and intuitive. At the critical level, the level of reflective, impartial moral thought, we can appeal only to Utilitarian principles. At the intuitive level we are allowed to make use of everyday moral rules and intuitions. Most of our ordinary life is lived on the intuitive level, but we attempt to approximate the critical level in our more reflective moments, in order to assess and correct our intuitive rules. In Affluent's situation, the thoroughgoing Harean would presumably engage in critical thinking to the best of her ability, as she has the time to do so. She would see clearly that the best act, from the critical perspective, would be to make a donation. She would also see that the best set of intuitive responses for her to adopt in her (non-critical) everyday life would include a disposition to favour her own children. If she were to make her current decision intuitively, Affluent would buy music lessons. An affluent Harean would thus appear to face a dilemma, as the critical and intuitive levels of thinking conflict. Hare claims to be able to integrate his two levels in practice in a way which is not too demanding. Sadly, he doesn't tell us how this is done. His most consistent response would seem to be as follows. The basis of morality is the Utilitarian principle of critical thinking. Intuitive thinking is justified only when critical thinking is not practical, and only insofar as it provides the best available approximation of the latter. In a situation where both critical thinking and intuitive thinking are possible, the agent should follow the former. This should not disrupt the agent's intuitive responses, as these are only brought into play when critical thinking is unavailable. This suggests that Hare cannot endorse Affluent's decision. Affluent should obey critical thinking rather than intuitive thinking. This procedure would be very demanding. So Hare will be unable to defeat the Demandingness Objection. (Hareans might reply that intuitive thinking is always to be used in everyday life, even when critical thinking is possible. However, this would undermine Hare's two-level moral theory. If critical thinking is never used in everyday life, when is it used? If we rule out Affluent's case, then presumably we must rule out every other situation in which an application of critical thinking might influence an agent's behaviour. Yet this would seem to render critical thinking entirely superfluous, as not even moral philosophers in their most reflective moments would be allowed to engage in it. This would certainly be very far from Hare's original intentions, as he himself indulges in critical thinking a great deal.)

alternative solutions, particularly ones which reject the controversial claim that Affluent's act is wrong.

In the remainder of Part ONE, I discuss theories which claim that Affluent's act is not wrong.

Chapter TWO: Collective Consequentialism

In this chapter I discuss Collective Consequentialist solutions to the Demandingness Objection.

Collective Consequentialists seek a less demanding, less impersonal moral theory than strict Individual Consequentialism. They do so, not by weakening the Consequentialist function, but by changing its mode of assessment. Instead of directly judging acts by their individual consequences, Collective Consequentialists assess acts in terms of the consequences of the collective acceptance of rules.

Collective Consequentialism has considerable intuitive appeal. It is natural to see moral philosophy as the search for the optimal set of moral rules, and to expect those rules to be those which produce the best consequences. Collective Consequentialism also seems to be fairer than Individual-Consequentialism, as it does not require an agent to do more than her "fair share" in promoting the good, even if others fail to do theirs. Given this initial plausibility, it's worth asking whether or not Collective Consequentialism provides a plausible response to the Demandingness Objection.

The most striking feature of Collective Consequentialism is its element of utopianism: acts are assessed by reference to an idealised world in which everyone does what they should. There are many varieties of Collective Consequentialism. They can all be characterised by the following simple formula:

The Utopian Act Formula : an act (x) is the right thing for an agent (A) to do in a choice situation (S) if and only if x would be performed, in S, by an inhabitant of the Ideal Society. Call that society Utopia, and call the relevant inhabitant A's moral role model, or

A*. Utopia is defined, relative to a set of possible societies, as the one which contains the most value.

There are several stages to this process:

- (1) Delineate the set of possible worlds from which Utopia must be chosen (call that set U);
- (2) Specify a function whose domain is the set U, and whose value will be Utopia;
- (3) Specify a function whose domain is a set of ordered pairs of agents and situations, and whose range is the set consisting of the inhabitants of Utopia. The value of this function will be A's moral role model with respect to that situation;
- (4) Specify a function whose domain is a set of ordered pairs of possible moral role models and possible situations, and whose range is the set of acts available to the agent in S. The value of this function will then be x (i.e., what the agent's moral role model would have done in that situation).

We shall deal with each of these in turn, focusing on a particular form of Collective Consequentialism, that of Brad Hooker. Hooker claims that, in contrast to Individual Consequentialism and other theories, Collective Consequentialism "calls for an amount of self-sacrifice that is not unreasonable".³⁴ Hooker's version of Collective Consequentialism, which he calls "Disposition/Rule Consequentialism (or rule-consequentialism for short)" is as follows:³⁵

³⁴Hooker 1990.

³⁵All subsequent references to Collective Consequentialism are to Hooker's particular theory, unless otherwise stated.

Hooker's CC : "an act is morally right if and only if it is called for by the set of desires and dispositions the having of which by everyone would result in at least as good consequences judged impartially as any other."

Hooker argues as follows. If each relatively well-off person contributed some relatively small percentage of his or her income to famine relief, there should be enough to feed the world. Perhaps ten percent from each of the well-off would be enough. Though giving that much may seem difficult enough for most of us, the demand that we do so does not seem unreasonable. Collective Consequentialism therefore seems to escape the Demandingness Objection.

I propose to accept that ten percent of their income, while significant, is not an unreasonable demand for many affluent people. However, I am still suspicious of Hooker's argument.

2.1: Formalising Collective Consequentialism

Before discussing specific objections to Collective Consequentialism, we must clarify exactly what the position involves. We do this by following Hooker's theory through each of the four stages outlined above.

2.1.1: Delineation of U

Obviously, much turns on the delineation of the set of possible societies from which our Utopia is chosen. A large number of features affect the amount of value in a society. These include:

- available technology, material resources, climate etc.
- existing institutions (political, social, religious, cultural etc.)
- characters, motives, values, beliefs of individuals.

In any Collective Consequentialism, some of these factors are fixed as they are in the actual world, while others are allowed to vary. For instance, we might say that our set of possible societies includes only those whose technology, resources and institutions are identical to ours, but that any combination of individual characters, motives or beliefs was allowed. Our set of possible societies would thus contain all combinations of individual characters, motives or beliefs possible in a society with our technology, resources and institutions. We could then describe Utopia as having the "ideal" combination of individual characters, motives or beliefs for such a society.

Collective Consequentialists tend to focus on those factors which are left variable, ignoring those which remain fixed. This is unfortunate, as there are usually links between the two. For instance, the pattern of moral characters in a society is a function, inter alia, of the social institutions prevalent in that society. Therefore, fixing the social institutions will limit the range of possible patterns of moral character. We cannot find our Utopia by asking two independent questions, "What are the current social institutions?" and "What is the ideal pattern of moral characters?", and then combine existing institutions with ideal pattern. That pattern of characters might not be one which could co-exist with those institutions. Rather, our second question must be dependent upon the first: "What is the ideal pattern of moral characters which could realistically have grown up alongside these institutions?"

Another problem is that the fixed factors may, in turn, be influenced by the variable ones. For instance, social institutions are not entirely inflexible. They are subject to alteration. If the pattern of moral characters in our society had been radically different from what it is,

this would undoubtedly have a significant affect on the structure of our institutions.

Therefore, in assessing our candidate societies, we must consider, not only what our institutions are, but what they might become. So even our fixed factors cannot be taken as cast in stone. This problem is discussed in detail in section 2.2.4 below.

Variable factors are thus constrained by fixed factors. Most Collective Consequentialists impose additional constraints. These may include:

The Societal Boundaries Thesis : Collective Consequentialists usually assume that the set of possible societies is the set of possibilities for a particular actual society, usually either an individual modern industrialised nation (Britain, USA), or the set of such nations taken as a whole. It is (implicitly) assumed that the division of the world *into* different societies remains much as it is. Other societies are usually ignored, or it is assumed that they remain more or less as they are.

For instance, Hooker doesn't want "everyone" to be confined to members of a particular society. He treats global problems in terms of a cooperating developed world and a (separate) cooperating non-developed world. The question is then: what would be the optimal response from the developed world as a whole? As we shall see in section 2.2.1, this is more plausible on some accounts of the causes of famine than on others.

The Homogeneity Thesis : Collective Consequentialists generally assume that the set of possible societies includes only societies in which (almost) everyone has the same moral character, or in which everyone follows the same moral rules. This Homogeneity Thesis clearly comes in many different degrees. At the extreme, it would assume that absolutely everyone must be exactly the same, in every respect, as everyone else. More realistic versions admit the presence of some people with different moral characters, some people

who don't follow the general rules. Also, the Homogeneity Thesis is often thought to be compatible with social institutions which assign different moral "roles" to different individuals. In section 2.2.6 I discuss the plausibility of this thesis, and the implications of rejecting it.³⁶

Limits on Possible Rules or Dispositions : For most Collective Consequentialists, the most important variable factor will be the moral dispositions of individuals, or the moral rules they follow. Different sets of rules or dispositions will be present in different possible societies. Yet, for any pattern of behaviour or any set of possible consequences, there will be infinitely many logically possible rules (or dispositions) which could have produced that pattern or set. If we're to pick out a unique Utopia, we need to rule out most of these possible sets. One popular constraint is the following:

The Teachability Constraint : only those rules or dispositions which could be taught to an ordinary human being are allowed. If we are keeping certain institutions fixed, we should allow only those rules or dispositions which those institutions would be capable of teaching to an ordinary human being. Also, the costs of teaching the rule or disposition (which will, presumably, increase in proportion to its complexity and its demandingness) must be borne in mind when assessing possible societies, especially if we are assuming that the total resources available within each society are more or less fixed.

Applying the Teachability Constraint will not be easy. It is discussed in detail (along with similar constraints) in section 2.2.4.

³⁶Hooker implicitly adopts the Homogeneity Thesis, with the exception of a Brandtian exception clause, which is discussed in detail in section 2.2.4.

2.1.2: Moving from U to Utopia

The second function is that which takes us from the set of possible worlds U to the single possible world Utopia. The simplest form of this function would be a maximising function. That is, one which designates the member of U which has the highest value as Utopia. Complexities arise if we move away from simple maximisation. The first complexity is the possibility of a draw or tie, where several different possible worlds have the same value (or perhaps incommensurable values). The most natural response in this situation will be to say that Utopia is the set of all such possible worlds.

A second complexity arises from the fact that, within any possible world, things which have value will be distributed over time. For instance, if a primary source of value is the lives of human beings, then the value of a possible world can be said to be distributed over time as different human beings will exist at different times. This raises the question of whether and to what extent Collective Consequentialists should discount the value of lives which exist at some time in the far future. It also raises the more serious question of the delineation of possible worlds. The most natural way to delineate possible worlds is to regard each possible world as a complete history of what happens at every point in time in the world, every consequence of every action being completely and exhaustively specified. The problem with this method is that our Utopia will then be a world in which every agent takes extraordinary risks which then miraculously pay off. So long as each risk has some chance of paying off, then there will be a possible world in which they all pay off, and this will be the world with the greatest total value.³⁷ This would have the undesirable consequence that Collective Consequentialism would tell agents to behave very riskily. The obvious solution to this dilemma is to adopt a less rigid delineation of possible worlds,

³⁷ Assuming, for simplicity's sake, that the risks are not somehow mutually exclusive. If they are, then the best possible world will be that with the greatest amount of compossible successful risk-taking.

regarding the possible world not as a complete history of everything that ever happens in the universe, but rather (at each time t) regarding the possible world as a completed history of what has happened in the world before t , plus the possibilities after t . The value of a given possible world would then be its expected value at time t (that is, the time at which the agent found herself in situation S).

A third complexity is the possibility of choosing a Utopia in terms of agent-relative rather than agent-neutral values. This would enable different agents to have different Utopias. Alternatively, we might adopt a hybrid function with agent-neutral values picking out a small sub-set of members of U which would serve as possible Utopias, with agent-relative values then being left to pick the one that best suited to the agent or her values.

A fourth complexity is the possibility of adopting a non-maximising function from U to Utopia. Two alternatives are particularly interesting. The first is the possibility of a satisficing function: a function which, instead of choosing the best possible world within U , picks out the set of possible worlds which are good enough. The agent could then emulate an inhabitant of any one of those good enough worlds. A Collective Consequentialism based on a satisficing function at this point would be less demanding and more permissive than one based on a maximising function.³⁸ A second type of non-maximising function would be a Kantian function. Like the maximiser, the Kantian will seek to pick the best possible world from set U . Unlike the maximiser, however, the Kantian does not equate "best" with "having the most value". This difference arises from the difference between the two conceptions of rationality which underlie the respective theories. For the maximising Collective Consequentialist, rationality is the maximisation of expected utility. For the Kantian, by contrast, to act rationally is to act on maxims which

³⁸The relationship between this form of satisficing and the more familiar Satisficing Consequentialism adopted by Michael Slote might be worth exploring. Slote's theory is discussed in detail in Chapter FOUR.

one can will as a universal law. The best possible world on this view will then be the possible world which it would be rational to choose from U, namely that world in which everyone acts on maxims which can be willed as universal laws.³⁹

2.1.3: Finding A's Moral Role Model

If, like Hooker, we endorse the Homogeneity Thesis, then, once we have found Utopia, we can find A's moral role model simply by picking any one of the inhabitants of Utopia. As these people are all effectively identical, it won't matter which one we choose. Of course, if we reject the Homogeneity Thesis, then a number of complexities arise. These are discussed in section 2.2.6 below.

2.1.4: What would A's moral role model do in S?

Once we've located A's moral role model (A*), we must ask what A* would do in situation S. Several factors may complicate this question:

Vagueness of Dispositions or Rules : in Collective Consequentialism, our variable factors include the moral dispositions of individuals, or the moral rules they follow. Sadly, such dispositions or rules may not fully determine an agent's actions. There may be no definite answer to the question: what would a person who had that disposition or followed that rule do in this situation?

Alien Situations : an agent's actions are most likely to become indeterminate in situations which are alien to the society in which the agent's dispositions were formed. For instance,

³⁹The relationship between a form of Collective Consequentialism which incorporates the Kantian function at this point, and traditional Kantian moral theories, might be worth exploring further.

if S bears little resemblance to A*'s daily life in Utopia, then A*'s dispositions or rules may fail to produce a definite response to situation S. I discuss the problem of alien situations in more depth in section 2.2.7.

A third complexity is the need for a precise specification of the phrase "A is in S". How much of what is essential for the description of my situation is tied to my being the particular person who I am, to my having the particular relationship which I have? What would it be for another person, specifically a Utopian, to be in precisely the same situation as me? A complete answer to these questions would take us well into the realms of complicated metaphysical questions of personal identity and the description of situations and events. Yet, without a complete specification, it is difficult to see how we can answer the question the Collective Consequentialist seeks to answer, namely: "How should I behave in my situation?"⁴⁰

A related possibility is that the statement "A is in S" may be contradictory. That is, it may be impossible for a Utopian to be in my situation. For instance, S might be a situation of having done something wrong, the situation of having made incompatible promises, or the situation of having made impossible or logically contradictory promises. But these are things which, by her very nature, a Utopian would not do. S will thus be a situation in which no Utopian could find herself.

Faced with this prospect, the Collective Consequentialist must say that when a proposition p is contradictory, propositions of the form, "If p were the case then q would be the case" are sometimes true and sometimes false. This is the only option which would enable her to say that some things are obligatory, permitted or forbidden, while other things are not.⁴¹

⁴⁰Philip Pettit and Michael Smith discuss related difficulties in Pettit and Smith 1993.

⁴¹To see this, consider the other two options available to the Collective Consequentialist. The first is to say that for any proposition p, if p is contradictory, then the proposition, "If p were the case then q would

However, this option presupposes a particular view about the nature of counter-possibles, which would need to be defended at this point.⁴² This problem is an extreme version of the problem of alien situations, which is discussed in section 2.2.7 below.

A final complexity arises if *x* is not an option for me; if, in my situation, a Utopian would do something which it is not open to me to do. For instance, *x* might be something which I cannot imagine, something which is psychologically impossible for me, or something which I lack the skills, techniques or knowledge to perform accurately. We could, of course, rule this possibility out by our definition of what it is for a Utopian to be in *S*. We might say, for instance, that part of what it is to be in my situation is to have available to you precisely the options which are available to me. However, this response threatens to undermine the whole Collective Consequentialist approach, as it may produce a theory in which what I should do is what I will do anyway. This is particularly true if we rule out options which are psychologically difficult or problematic for me to perform.⁴³ It seems then that Collective Consequentialists should accept the possibility that what a Utopian would do if she were in my situation is something which is not a genuine option for me. However, this raises the possibility of my being under an effectively impossible obligation.

be the case" is necessarily true. In this case, we would have to say that, if it is impossible for my moral role model to be in my situation, then claims of the form "If a Utopian were in situation *S* then they would do *x*" will be true for any *x*. In other words, everything will be obligatory. Also, for any *x*, a claim of the form "If a Utopian were in situation *S* then they would not do *x*" would also be true. So all actions will be forbidden as well as obligatory. While we might want to admit that there are some situations in which more than one action is obligatory, or in which all possible actions are forbidden, it seems absurd to suggest that every single situation is one in which every single action is both obligatory and forbidden. This first option is not viable. The second option is to say that for any proposition *p*, if *p* is contradictory, then any proposition of the form, "If *p* were the case then *q* would be the case" is necessarily false. On this view, any statement of the form "If a Utopian were in situation *S* then they would not do *x*" will be false for any *x*, so every action will be permitted. Again, while it may be acceptable to say that there are some situations in which more than one action is permitted, it seems absurd to say that every situation is one in which every action is permitted. So the second options is not acceptable either.

⁴²As the impossibility of my role model being in my situation most naturally arises when my situation is one of my having done something wrong, it might be profitable to link the present discussion to discussions of conditional obligation in deontic logic. (On which see, for instance, Chisholm 1963, DeCew 1981, Greenspan 1975, Lewis 1973, Mott 1973, and von Wright 1969.)

⁴³Also, this response greatly increases the chance that it will be impossible for a Utopian to be in my situation. If being in my situation means only recognising the options that I am able to recognise, and if, by her very nature, a Utopian is someone whose ability to recognise options is greater than my own, then, by definition, she cannot ever be in my situation.

If x is what I should do, and x is not a genuine option for me, then what I should do is something which in a very real sense I cannot do.

2.1.5: Some Short-cuts

The procedure for finding x looks very daunting. Fortunately, we don't always have to go through all the steps thoroughly. The Collective Consequentialist's primary interest is in locating x , in finding out what is the right thing to do in our situation. We're not interested in Utopia for its own sake. Therefore, if we can find x without exactly defining Utopia, we will be satisfied.

Assume, for instance, that instead of a single Utopia, we have a set of possible societies, which we're unable to choose between. Call that set $U+$. Now, for any particular member of $U+$ (call it U_i), we'll have a counterfactual of the following form: "If an inhabitant of U_i were in S , she'd do x_i ". Call the set of all the x_i , X . Now, X will be the set of possible right things. In the simplest case, X will be a unitary set. That is, all the x_i will be identical. We can then identify x with the solitary member of X , thus getting our right thing without definitely specifying our Utopia. Sometimes, however, X will not be unitary: different possible Utopians will do different things. We must distinguish three separate situations: uncertainty, indeterminacy and dead-heats. In the first, we know that there is a definite answer to the question posed by the Utopian Act Formula, but we're unable to find it. We know that some members of our set of societies are better than others, but we lack the information to distinguish them. In this case, if we adhere to the Utopian Act Formula, we must conclude that, while there is a definite right thing to do in this situation, we don't know what it is (though we may know what it isn't). In cases of genuine indeterminacy, we accept that there is no fact of the matter as to which member of our set of possible societies contains the most value. We won't know that a particular option is obligatory, but

we will probably be able to find several which are permitted. In cases of dead-heat, we have either two or more possible societies with exactly the same value.

If X is non-unitary, then it is most natural to say that all (and only) the members of X are permitted. We can thus reformulate our Utopian Act Formula principle as follows:

The Revised Utopian Act Formula : an act (x) is permitted for an agent (A) in a choice situation (S) if and only if x would have been performed, in S , by at least one possible moral role model of A 's (i.e.: by at least one inhabitant of at least one member of $U+$, as defined above).

Collective Consequentialists often talk as if it were necessary to have a theory which marked off and located a definite "right thing to do" in any particular situation. This demand is unreasonable. In many situations there will be no precisely delineated set of obligatory (or even permissible) actions, while at other times we won't be able to find that set. This doesn't mean Collective Consequentialism provides no useful guide to action. We can say to an agent: you must do something which is permitted. These actions are definitely permitted, these others are definitely not permitted, while those over there may or may not be permitted. This is probably the most we can ask of any moral theory.

2.2: Objections to Collective Consequentialism

We now turn to my objections to Collective Consequentialism. These fall into two groups. The first group consists of objections to the Collective Consequentialist's response to the Demandingness Objection in particular (sections 2.2.1 through 2.2.4). The second group consists of more general objections to Collective Consequentialism (sections 2.2.5 through 2.2.7).

2.2.1: The Wrong Facts Objection

Recall our original story:

Affluent's Tale : Affluent is a well-off member of a contemporary first world society, who already makes significant donations to charity. Affluent is sitting at her desk with her cheque book. In front of her are two pamphlets: one from a reputable international aid organisation, the other from her children's prospective music teacher. Affluent has enough money to either buy the music lessons or make a donation to the charity, but not both. Because of her love for her children, she buys the lessons, even though she knows that the money would have done far more good if sent to the charity.

Affluent knows that others will not act optimally. Therefore, under Collective Consequentialism, the degree of sacrifice which is required of her will be the level of sacrifice which would be optimal under full compliance. The demands which morality places on Affluent will thus depend upon the following question: What would be the optimal response to famine from the developed world as a whole, if everyone did their bit?

Unfortunately, this is a terribly difficult question, which cannot be answered without a credible and comprehensive account of the causes of famine and poverty. There are three traditional analyses: Developmentalist (famine is caused by a sudden decline in the food supply); Neo-Malthusian (overpopulation causes famine); and Micro-Economic (famine is caused by the collapse of direct or exchange-entitlement economic relationships in the local situation).

Jean Dreze and Amartya Sen argue for the third analysis of the causes of famine.⁴⁴ They argue that the severity of famine depends largely upon local conditions and institutions, in particular political institutions and public relief programmes. Famine, they argue, is less dependent on the availability of food or on the population. In their analysis, the behaviour of developing countries, the countries where famine is a possibility, is in many ways far more important than the behaviour of the developed world in terms of giving aid or assistance.⁴⁵

George Lucas has argued that,⁴⁶ if we accept this Micro-Economic analysis, our response to famine should be something like the following:

- (1) local governments should provide "an elaborate economic safety net";
- (2) developed countries should provide monetary assistance to those local governments;
- (3) in the absence of safety nets, developed countries could establish them directly in the affected countries.

How much individual sacrifice would be involved in such comprehensive aid programs? At first glance, it looks as if dispositions such as: "vote for governments which will institute fair and effective aid programs" and "pay your taxes" will be sufficient. Yet such dispositions require hardly any sacrifice at all.⁴⁷

⁴⁴Dreze and Sen 1989.

⁴⁵Note that this analysis of famine would present problems for Hooker's decision to treat famine in terms of the reaction of the developed world to a set-pattern of developing-world behaviour.

⁴⁶Lucas 1990.

⁴⁷Such empirical complexities are discussed further in sections 2.2.3 and 2.2.5.

For our present purposes, we can bypass these empirical complexities. This is because, irrespective of whether or not these questions can be answered, the very fact that Hooker requires us to ask them constitutes an objection to his theory. Collective Consequentialism requires us to be experts in global political economy to an extraordinary degree in order to decide how much, or indeed whether, we should donate to Oxfam. In situations where compliance with the best global response will be extremely partial, it is hard to see why such questions should be so relevant. Hooker makes the answer to the question "What should we do here and now?" depend on precisely the wrong empirical questions.

We can see this by discussing three sets of possible variations on Affluent's story.⁴⁸ In the first set of experiments, we vary the number of starving people. Consider three possible variations on Affluent's Tale (each exactly like the original except where stated):

Few Poor : there are only 1 million people starving in an affluent world.

Medium Poor : there are 50 million people starving or facing famine.

Many Poor : famine breaks out across China in the next century. There are 2,500 million people starving.

In all three cases, the eradication of famine lies far beyond Affluent's means. All she is able to do is make a very small dent on the overall problem.

Under Collective Consequentialism, the answer to the question "How much should Affluent donate to Oxfam?" will be radically different in each of the three cases. For instance, let us make the crude assumption that, in Medium Poor, the optimal global

⁴⁸An earlier version of the remainder of this objection appeared as Mulgan 1994.

response to poverty would be a donation equivalent to 10% of the income of all inhabitants of the developed world. Things will be better overall if everyone gives 10% than if everyone gives either more or less than that amount. So Affluent should donate 10% in Medium Poor. In Few Poor the appropriate sacrifice will be one hundredth of 10% of Affluent's income. The reason for this is simply that if 10% of the incomes of all the affluent people will save 50 million people from starvation, then roughly one fiftieth of that amount will be required to save 1 million people from starvation. At the other extreme, if 10% of the income of the affluent people will save 50 million people from starvation, then the saving of 2,500 million people would seem to require virtually all, if not all, of the income of those in developed countries. In Many Poor, then, Affluent would be required to donate virtually all of her income to Oxfam.

These extreme differences seem unreasonable. In Few Poor, it seems ridiculous for Affluent to only donate one fifth of 1% of her income, simply because there are *only* 1 million people starving, rather than 50 million. Similarly, it seems unreasonable to demand of the Many Poor Affluent that she give up all of her income simply because there are 2,500 million people starving rather than *only* 50 million. The difference between the situations does not seem sufficient to justify the enormous difference in sacrifice required of the three Affluents. In order to see this, we can unite the three cases in a single thought experiment:

Affluent's First Ignorance : Affluent does not know how many people are starving. Indeed, she is so radically misinformed that all she knows is that she is in one of the three situations described (Few Poor, Medium Poor, or Many Poor). As she sits down at her desk, Affluent is able to find out exactly what situation she is in. (Perhaps by ploughing laboriously through the fine print of the latest World Bank report.)

Does Affluent really need to do this before deciding how much money to give to Oxfam? Collective Consequentialism says that she does. Until she has resolved her ignorance, Affluent doesn't know whether to donate 0.2%, 10% or (nearly) all of her income. Yet this seems implausible. If a donation of 10% is the appropriate sacrifice if there are 50 million people starving, then giving something like 10% seems to be the right response, whether there are actually 50 million or only 1 million or as many as 2,500 million. Once we know that millions of people are starving, the exact number of millions does not seem to bear the weight which is accorded it in Collective Consequentialism.

My objection here is not merely that Collective Consequentialism makes Affluent's obligations dependent on specific empirical facts. After all, Individual Consequentialism requires similarly detailed information, as what an agent should do depends on the precise consequences of her actions. However, unlike Individual Consequentialism, Collective Consequentialism requires extensive knowledge of facts which have no effect on the consequences of our actions. Affluent knows that her most efficient aid-contribution method is to write a cheque to Oxfam. She also knows (roughly) how much good would be produced by any given level of donation. Does she really need to know how many other people will fail to make adequate donations?

In our second thought experiment, we vary not the number of starving people, but rather the number of people in the developed world. This produces the follow tale:

Affluent's Second Ignorance : Affluent, as she sits at her desk, is drastically misinformed about the size of the developed world. She knows that the number of people in the developed world is very large, but she doesn't know exactly how large it is. She knows that she's in one of the following three situations:

Few Rich : the developed world is very small, containing only 1 million people.

Medium Rich : the developed world includes 50 million people.

Many Rich : Due to rapid economic progress in China, the developed world contains 2,500 million people.

Under Collective Consequentialism, Affluent must resolve her ignorance before she has any idea at all how much she is required to give to Oxfam. Assume that in Medium Rich Affluent should donate 10% of her income. If 10% of the income of 50 million people is enough to remove poverty, then one fifth of 1% of the income of 2,500 million people should also be sufficient. Therefore, in Many Rich Affluent is required to donate only one fifth of 1% of her income. By contrast, if there are only 1 million people in the developed world, they will have to sacrifice all or virtually all of their income and resources. In Few Rich the sacrifice required of Affluent is extreme.

This seems unreasonable. If 10% is enough if there turn out to be 50 million people in the developed world, why shouldn't it also be enough if there turn out to be only 1 million, or on the other hand, not too much even if there turn out to be 2,500 million people? Given that there are many millions of people in the developed world, and that the vast majority of them are not doing what they should, the precise number of millions who are not doing their share does not seem to be as important to Affluent's obligations as Collective Consequentialism requires it to be.

Collective Consequentialism falls into these mistakes because it presupposes a rigid relationship between the sacrifice required of any particular individual and certain features of their global situation. Crudely, the sacrifice which Collective Consequentialism requires

of an individual is equal to the total sacrifice required of all moral agents in that situation divided by the number of agents able to make the appropriate sacrifice.

In our third set of stories, we vary not the number of people facing starvation, nor the number of people in the developed world, but rather the nature of political economy: in particular, the nature of the laws that govern the existence of famine. Consider three possibilities:

Affluent's Third Ignorance : Affluent is completely ignorant of political economy. She knows only that one of the following is an accurate description of the nature of her world:

Developmentland : a world in which the Developmentalist account holds true; in which famine is caused by a sudden decline in the food supply.

Malthusia : the Neo-Malthusian view that over-population is the primary cause of famine is correct.

Senland : Sen's micro-economic analysis in which famine is caused largely by the collapse of direct or exchange entitlement economic conditions is the correct analysis.

As Lucas has argued, the appropriate global response to famine depends crucially on which one of these three worlds is the actual one. Sacrifices under full compliance vary greatly between the three worlds. So, under Collective Consequentialism, Affluent must resolve her doubt before she acts. Before she knows how much to donate, Affluent must become an expert on global political economy.

This seems unreasonable. Affluent's behaviour has little chance of bringing about the optimal global response, whatever it may happen to be. Her questions should only be "Will giving money to Oxfam do good as things are? How much good will it do? Therefore, how much am I required to give, given other considerations?" The nature of the precise optimal global response does not seem to be one of those other considerations which Affluent must address. If 10% is about right on one economic account, why shouldn't it be right on other economic accounts, as the amount of good which that particular donation will do would be more or less the same in each of the three cases?

I do not claim that these empirical differences should have no effect at all on Affluent's obligations. I claim only that they should not affect her obligations to the extreme extent which Hooker says that they do.

Faced with this barrage of examples, Hooker has a number of replies open to him. His actual response is as follows.⁴⁹ Collective Consequentialism incorporates a Brandtian Complexity Constraint: the notion of "teaching costs". Our choice of moral rules is limited to those which could be taught to a human population at non-prohibitive social cost. A rule requiring Affluent to donate virtually all her income to charity could not be effectively taught. Therefore, Collective Consequentialism cannot demand so much. Its demands will not be subject to the wild fluctuations alleged in the Wrong Facts Objection.

Hooker's Complexity Constraint is discussed in detail in section 2.2.4. At this point, I wish to suggest that, whatever its fate overall, the constraint is unable to provide a response to the Wrong Facts Objection. For one thing, it is by no means obvious that, in certain circumstances, a significantly demanding morality could not be taught. In particular, consider a case such as Many Poor. The benefits of inculcating a demanding morality

⁴⁹Hooker 1994a, pp. 24-27.

among the developed world of Many Poor would be immense. If it were possible to do so, any social or psychological costs would presumably be outweighed by those benefits. Perhaps the resulting morality would not be one which would tell Affluent to donate all of her income. However, it probably would tell her to donate far more than 10%. The donation required in Many Poor would thus be far greater than that required in Medium Poor. The discrepancy to which the Wrong Facts Objection objects would remain.

A second problem with this response is that it seeks to defeat the Wrong Facts Objection by showing that Collective Consequentialism is not too demanding. Yet the Wrong Facts Objection claims that, in some cases, Collective Consequentialism does not demand enough. Hooker's reply leaves this half of the objection untouched. Take the case of Few Poor, where we objected that Affluent would only have to donate 0.2% of her income. A rule requiring such donations could clearly be taught without prohibitive costs. So the Complexity Constraint doesn't rule it out.

Hooker might reply that the optimal set of rules will always include a certain basic level of benevolence, which can be inculcated at negligible cost. However, he would then face a dilemma. Either this basic level of benevolence translates to a donation of approximately 10% in Affluent's original situation, or it does not. If it does not, then the disparity alleged by the Wrong Facts Objection remains. If it does, then Hooker is asking us to believe that there is little significant difference between the demands of a principle of benevolence which can be inculcated at negligible cost, and those of a principle which is too demanding to be taught at all. This would be a very contentious claim.

In any event, even if Hooker is able to get the right answer with this response, he will get it for the wrong reasons. The Wrong Facts Objection alleges that Collective Consequentialism takes too much account of (morally) irrelevant facts. Hooker's reply is

not that he does not take account of such facts, but merely that they do not greatly affect Affluent's obligations. This is not enough. The fact remains that, if a significant range of principles of benevolence were teachable, then Collective Consequentialism would make vastly different demands on Affluent in different situations. This suggests that this approach is misguided.

2.2.2: Against Brandtian Exception Clauses

In a recent paper,⁵⁰ Hooker defines Rule Consequentialism as follows:

An act is morally permissible if, and only if, it is allowed by the code of rules whose collective acceptance could reasonably be expected to produce the best consequences as judged from an impartial point of view.

He then goes on to say that:

by collective acceptance, I mean acceptance by everyone except young children, the mentally impaired, and those whose acceptance is precluded by a description of some moral problem situations.

The last clause is adapted from Richard Brandt,⁵¹ and as Hooker explains it is "made because we need rules for dealing with those who misbehave and for dealing with problems created by the non-compliance of others". I shall now argue that Brandt's exception clause is unacceptable as it causes his theory to collapse into Individual

⁵⁰Hooker 1994a.

⁵¹See Brandt 1992a, 1992b and 1992c.

Consequentialism in cases such as famine, which result from (near) universal failures to act.

To begin with, consider four simple cases:

Famine Just Happens : through no action or inaction or fault of any particular person or group of people some people face starvation or famine. For instance, a large and adequate store of grain may have been destroyed by an unpreventable (or unforeseeable) act of God

The Brigands' Famine : there is plenty of grain in the country, but the soldiers of one side in a civil war burn all the crops in the territory of the other side.

The Hoarders' Famine : famine arises because those who own the grain decide to hoard it. They refuse to sell it or give it to those who require it (or, at least, they refuse to sell it to the latter at a price which they can possibly afford).

The Ignorers' Famine : famine results (in part) from the collective inactivity, or inaction of developed countries.

In Famine Just Happens, we don't have to exclude anyone from the domain of collective acceptance, as no one's wrong-doing or action brings about the situation which is problematic. In the Brigands' Famine we must exclude from our domain of collective acceptance the people who are responsible for burning or destroying the grain. This is because their collective acceptance of desirable moral rules would have been inconsistent with the existence of the problem under discussion. In the Hoarders' Famine, the problem we're discussing would not arise without the behaviour of the grain hoarders. Therefore, we must look at rules whose collective acceptance includes everyone except the grain

hoarders. Note that the grain hoarders cause the famine as a result of inaction. The grain is not destroyed. Nothing is done to it. It is just as a result of inaction prevented from getting to the people who need it. We are excluding people from the domain of collective acceptance on the basis of what they do not do, as well as what they do.

Our final tale, the Ignorers' Famine, is the case Hooker describes. An appropriate response from developed countries would remove the famine. The behaviour of ordinary citizens in the developed West is necessary for the continued existence of the problem under discussion. Therefore, they must be excluded from the domain of collective acceptance.

But now our Collective Consequentialism is hardly collective at all, for to decide what I should do, I must look at all possible sets of rules and examine their collective acceptance by everybody except brigands, grain hoarders, and citizens of the developed West. Our universal acceptance is now a (near) universal rejection.⁵²

Excluding virtually everyone in the developed West greatly reduces the appeal of Collective Consequentialism. In particular, it weakens Hooker's claim that his theory is less demanding than Individual Consequentialism. When deciding how much money I should give to Oxfam, I should ask what amount of money would be sufficient if given by all

⁵²Note that if we exclude all those whose inaction is a necessary part of describing the problem situation, then we may end up with something like Donald Regan's Cooperative Consequentialism (see Regan 1980). Regan's Cooperative Consequentialism is fairly complicated, but it consists essentially of recommending the following decision procedure to moral agents.

Step 1: identify all potential cooperators.

Step 2: predict the likely responses of non-cooperators to various patterns of behaviour by potential cooperators. Treat those responses as given.

Step 3: play my role in the optimum cooperative strategy.

Such a procedure is almost identical to that which would be followed by someone seeking to find the appropriate Collective Consequentialist response to a disaster, if they were factoring out the behaviour of those whose actions contributed to that disaster (that is, the non-cooperators). We should also note that, in practice, Regan's Cooperative Consequentialism seems perilously close to an Individual Consequentialism which cautions me to take account of the effects my actions are likely to have on potentially sympathetic others. A full discussion of Regan's theory is beyond the scope of this chapter. However, the possible collapse from Hooker's Collective Consequentialism through Regan's Cooperative Consequentialism to traditional Individual Consequentialism serves to highlight the precarious nature of the Collective Consequentialist's search for an acceptable response to problems caused by non-compliance.

those people in the West who are already giving significant amounts to Oxfam. Clearly, my share of the burden of relieving famine will be far greater in this case than in Hooker's original version.

It may seem possible for Collective Consequentialists to avoid this objection by drawing a distinction between action and inaction. However reasonable it may seem, this move is not really available to the Consequentialist. A major plank of the Consequentialist's position has traditionally been the claim that there is no moral difference between doing and allowing. A version of Collective Consequentialism which survives only by smuggling such a distinction into a Brandtian exception clause would be unacceptable.⁵³

Another problem with an over-exclusionary Brandtian exception clause is that it would require different sets of moral rules for almost every particular problem situation. Many current problems only exist or persist because of the inaction of some group of people. For instance, environmental problems are often exacerbated by the inaction or ignorance of the majority of people in the developed West. So those people would need to be excluded from the domain of collective acceptance.⁵⁴

Collective Consequentialists may reply with an appeal to the distinction between following a rule and accepting a rule.⁵⁵ They may argue that, in a case such as the Ignorers' Famine, the majority of people accept a moral rule which, if everyone followed it, would alleviate the problem. They simply don't follow that rule. However, this move would seem to strain the notion of "accepting" a moral rule. Do we really want to say that a rule may be accepted

⁵³See also the discussion of Kagan's analogous objection to Scheffler in Chapter FIVE.

⁵⁴Similarly, if our problem was poverty in the UK, we might need to exclude from the domain of collective acceptance all those people who are in the habit of voting for the current Government, as the combination of their action in voting that Government in and their collective inaction in not providing sufficient charitable donations may be a significant cause of the existing poverty.

⁵⁵This distinction is discussed in more detail in section 2.2.4 below.

within a given community even if almost no one follows it? Also, this response would be incompatible with Hooker's strategy in two key ways. First, if a rule can be accepted without being followed, then we will have little reason to accept Hooker's claim that very demanding rules cannot be accepted. Such rules will be demanding to follow, but not to accept. Second, Hooker's theory gives great weight to the consequences of a rule's being accepted. If a rule can be accepted without being followed, then it's not clear either what those consequences will be, or why we should be particularly interested in them. The significance of the acceptance of a rule presumably lies in the fact that that acceptance makes it more likely (though not inevitable) that the rule will be followed.

2.2.3: Is Collective Consequentialism Undemanding?

Let us grant Hooker's assumption that Utopia will be a place where everyone gives 10% of their income to charity or to some other means of alleviating poverty or famine, and that, as a result, there is no poverty or famine in Utopia. We must then ask firstly, "What is the rule that these people are following?", and secondly, "How would someone following that rule behave in the situation in which Affluent finds herself?" In particular, what would that Utopian rule say about situations of extremely partial compliance?

To give a very brief idea of the complexities involved in this question, consider just four possible rules, all of which might produce the behaviour Hooker expects in Utopia.

Rule A : give 10% to the best available charity, irrespective of what other people do.

Rule B : give 10% of your income to the best available charity if everyone else or enough other people will do so, otherwise give nothing, as any contribution you made would not be sufficient to remove the problem of poverty.

Rule C : give 10% of your income to charity if enough others will do so, otherwise give more on the grounds that there will be more suffering for you to alleviate in that case.⁵⁶

Rule D : follow the optimal set of rules (those which are current in Utopia) if enough others will follow those rules, otherwise maximise utility as best you can.

Hooker's discussion seems to assume that we're talking about a Utopia in which the rule governing people's behaviour is Rule A. Even in Affluent's situation of almost total non-compliance, Hooker's Utopian will only give 10% of her income, as that is exactly what she would have given in Utopia, or at least in a situation in which everyone else had given the same amount. However, it seems unlikely that the Utopian rule would be as rigid as Rule A.⁵⁷

However, once we admit the need for flexibility, the Collective Consequentialist seems to be on a slippery slope leading to Rule D, which gives the same recommendations as Individual Act Utilitarianism. We need some reason to stop between A and D. Presumably we want some form of Rule C which is not too demanding.

There are two problems here. The first is to specify such a Utopian Rule. This problem is discussed in section 2.2.2 above and in sections 2.2.4 and 2.2.7 below. The second is to guarantee that the resulting rule is not unduly demanding in everyday situations. It is to the latter problem that we now turn.

⁵⁶There are many possible variations of Rule C, depending on precisely how much more one would give if others fail to follow this rule.

⁵⁷Indeed, Hooker needs some such flexibility if he is to answer the objections raised in sections 2.2.2 above and 2.2.4 below.

Hooker considers only the amount of money required to lift everyone in the world above the poverty line. He implicitly assumes that further redistribution would not yield further improvements in welfare. This seems unwarranted. Assume, for the sake of argument, that 10% of the wealth of the developed world would be just enough to raise everyone in the world above the poverty line. This redistribution would leave the millions of people who now live in poverty living in almost-poverty, while the affluent few would still be living in almost-affluence. It seems reasonable to assume that a further redistribution which brought the almost-poor further above the poverty line, while still leaving them worse off than the more affluent, would provide an additional increase in welfare.

If we assume that there are enough resources in the world to provide at least an adequate life for everyone, and we also assume that, other things being equal, any given sum of money provides a greater amount of welfare for a poorer person than for a wealthier one, then we might conclude that, in Utopia, the developed nations would aim for perfect redistribution of the world's resources. It is difficult to estimate the effect of perfect redistribution. However, one thing seems clear: perfect redistribution would require a much greater sacrifice from well-off people in affluent societies than Hooker's token 10%.

I see several objections to this line of argument. The most interesting is the following: perfect redistribution would be inefficient, and require sacrifices which run contrary to human nature. This is an analogue of the familiar argument that extensive tax-benefit programs reduce people's incentive to work, as the benefits of working as compared with not working are greatly reduced. Perfect redistribution is an extreme form of tax-benefit program, where everyone receives exactly the same standard of living, irrespective of how hard (or even whether) they work. It is argued that, in such a system, entrepreneurs, producers and workers would not take the risks or make the sacrifices which are necessary for economic production. The amount of wealth available to be distributed would thus fall

dramatically. This fall would be so great that perfect redistribution would reduce aggregate welfare.

There is certainly something in this argument. Perfect redistribution is unrealistic. However, it's important to tease out the precise nature of this unrealism. Collective Consequentialism is a counterfactual moral theory (that is, a theory which decides what is right by asking what would be the case in some hypothetical situation, such as full compliance). Any such theory is built around unrealistic situations. The ideal society is not the actual world. The fact that a particular scenario is "unrealistic" does not necessarily rule it out as a potential ideal world. The crucial question is whether or not it's unrealistic in the wrong way.

The usual presentation of the taxes reduce incentives argument presents perfect redistribution as "economically impossible". The relationships between effective marginal tax rates, overall tax burden and productivity are said to be governed by "economic laws", much as the relationships between mass, acceleration and force are governed by physical laws. In morality, we don't usually decide what to do on the basis of what we might have done if the laws of physics had been radically different. Similarly, we shouldn't decide what to do by considering a situation in which the laws of economics are radically altered.

I am not convinced of the analogy between economic laws and physical laws. Economic laws come in two types: those dealing with physical relationships (e.g., laws of increasing returns to scale or diminishing marginal productivity of capital), and those dealing with human behaviour (e.g., laws describing people's response to certain changes in their incentive structures). I agree that the former class of economic laws should not be varied in moral deliberation, anymore than should the laws of physics.⁵⁸ However, I do not believe

⁵⁸Except in so far as appropriate research might alter even the physical laws of production.

that all psychological "laws" concerning human behaviour should necessarily be held constant within a theory of moral counterfactuals. Such a theory is primarily concerned with human behaviour. If we're not allowed to vary any generalisation regarding such behaviour, then our moral role models will be carbon copies of ourselves. Our theory will thus be effectively unable to offer us any suggestions for improvement. It will never be able to say that we have done the wrong thing. This seems unreasonably restrictive.

Economic laws regarding the disincentive effects of perfect redistribution are, at least partially, psychological. It is thus reasonable to ask whether or not these laws should remain fixed in our search for an ideal society against which to measure our behaviour. The following fantasy may prove instructive in discussing this question:⁵⁹

The Centralised Charity : Shareland is a world with resources, population and technology approximately equal to our own. The economy of Shareland is capitalist: resources are allocated by market forces rather than by centralised planning. Firms and means of production are privately owned and operated. Initially, everyone receives only their market income. At the end of each month, each inhabitant of Shareland fills out a form saying how much market income they received that month. Everyone fills out their form honestly and accurately. These forms are then sent to a central organisation, which calculates the average market income over all inhabitants. This average is then published. People who received more than the average pay their "extra" income into a central account, which is then used to top-up the incomes of those whose incomes were below average. There is no negative effect on production, as the system is entirely voluntary and Sharelanders are motivated by a desire to equalise resources.

⁵⁹A similar tale is discussed in Carens 1981.

This story is certainly unrealistic. But it is not unrealistic in the way that, for instance, a story in which rocks were edible would be unrealistic. The Centralised Charity makes no unrealistic assumptions regarding the physical limits of various factors of production, or regarding the efficiency of technological processes. Its unrealism is purely psychological: the Sharelanders system wouldn't work, not because it couldn't work, but because people wouldn't let it work. People would fill out false returns, claim more than their "fair" (in Shareland terms) share, misuse the centralised store of information required to set up the system, etc.⁶⁰

The Shareland system is unrealistic primarily because people wouldn't behave in the totally selfless and scrupulously honest manner which it requires. However, one of the central questions which Collective Consequentialists are trying to answer is the following: How selfless can we reasonably expect people to be? Obviously, if we rule out any possible world in which people are more selfless than they would ordinarily be, we will inevitably end up with the answer that we can't reasonably expect people to be any less selfless than they are. This may be the right answer, but even if it is we shouldn't arrive at it quite so easily. Surely we must allow for the possibility that the ideal society is one in which people are more selfless than we would be in similar circumstances.

So we can't rule out Shareland simply because its inhabitants are more selfless than we are (or than we would be). However, we might rule it out by means of a stronger claim. For instance, we might suggest, not just that we wouldn't be as selfless as the Sharelanders, but also that we couldn't. We might even go so far as to claim that no human beings could ever be completely selfish, irrespective of how well they'd been brought up. The

⁶⁰Of course, in a society in which it was established, the Shareland system would not be so easy to subvert. If everyone had exactly the same wealth, it would be very difficult to do anything with any ill-gotten gains one might acquire. Any conspicuous wealth would be suspicious. The problem with the Shareland system is that it couldn't get off the ground in the first place. Also, mass corruption might arise all at once, making each person's chances of being punished greatly reduced.

Sharelanders aren't ideal people, because they're not people at all. A less extreme claim would be that, although people could be made as selfless as Sharelanders, the costs of doing so (in psychological or other terms) would outweigh the benefits.

This line of argument is more promising. If a certain way of life is really impossible for human beings, then we may well rule it out as a stand-point from which to assess our moral behaviour. However, it's obviously very difficult to say exactly which (logically) possible human societies are humanly possible. We need a precise account of the boundaries of human nature, which we don't have.

However, even if we decide that the original version of our story is unrealistic, we can still use the Centralised Charity as a conceptual tool. We can progressively water it down until we find something which is acceptable: the redistribution carried out by the system need not be total, safeguards can be introduced to prevent false returns, abuse of information, etc. Eventually, we would end up with something akin to the tax-benefit systems currently operated in modern welfare states. The fact that these systems exist suggests that they are humanly possible. We can conclude that, at some point, the Centralised Charity becomes an acceptable story.

Of course, it may not be possible to say exactly when this point occurs. We won't be able to specify exactly how much redistribution is involved in the most redistributive plausible version of the Centralised Charity. Also, unlike the original system, the redistribution mechanisms in watered-down stories may involve inefficiencies: if people are no longer perfectly selfless, redistribution may affect their incentives. So we can't assume that the best acceptable version of the Centralised Charity is the one with the greatest degree of redistribution.

We could rank all possible versions of the Centralised Charity along two scales: degree of redistribution and degree of inefficiency. Using these two measures, we could (in theory) generate a single measure of aggregate welfare, which will be increased by redistribution but reduced by inefficiency. In the set of acceptable versions of our story, an increase in redistribution will tend to be accompanied by an increase in inefficiency. So the net effect on aggregate welfare of a given change in redistribution can't be determined a priori.

Hooker's claim, in effect, is that, in the class of acceptable versions of the Centralised Charity, aggregate welfare is maximised when redistribution removes only 10% of the income of well-off people in affluent societies. This claim, as we have seen, will be very difficult to assess. However, Hooker's figure does, on the face of it, seem rather low. After all, 10% is a far lower figure than the rate of taxation in many developed nations, where redistribution doesn't seem to have crippling disincentive effects.^{61 62}

All these questions are incredibly complicated. Until we've gone at least some way towards answering them, we can't simply conclude that Collective Consequentialism really is significantly less demanding than Individual Consequentialism.

⁶¹Hooker may reply, of course, that his 10% is on top of current taxation, but even then the total disincentive need not be crippling.

⁶²Also, recall the conclusion we reached in Chapter ONE (in the discussion of blameless wrongdoing). I suggested there that, on any acceptable moral theory, the limits of moral obligation must be distinct from the limits of psychological possibility (i.e., there must be things which are psychologically possible but not morally required). It appears that Collective Consequentialism violates this desideratum. See also section 2.2.4 below.

2.2.4: Brandtian Complexity Constraints

One common objection to Collective Consequentialism is the following:

The Partial Compliance Objection : Because it chooses moral practices or moral rules on the basis of what would happen if everybody complied with those rules or followed those practices, Collective Consequentialism gives undesirable results in situations of partial compliance where not everyone conforms with the rule in question.⁶³ For instance, assume that I am living in Sweden prior to 1967. I decide that the best rule for everyone to follow would be the rule of driving on the right-hand side of the road. As a Collective Consequentialist, I drive on the right-hand side of the road, despite the fact that everyone else is driving on the left. The results are not pleasant.

To avoid this objection, Collective Consequentialists usually say that their rules include clauses of the following form: Do x, unless doing x will lead to great disaster because everyone else is not doing x, in which case do y (where y is something which will not have disastrous results). So, in the Sweden example, the rule would be: Drive on the right unless driving on the right will have disastrous consequences because everyone else is driving on the left, in which case drive on the left.

The introduction of such complexity into the rules however, leads to another objection, due to David Lyons:⁶⁴

⁶³Collective Consequentialism is best framed in terms of acceptance rather than compliance, so this objection should perhaps be phrased in terms of "Partial Acceptance". As the term "Partial Compliance" is more familiar, I continue to use it in the text.

⁶⁴Lyons 1965.

The Co-extensionality Objection : Collective Consequentialism collapses into Individual Consequentialism. The rules it will be best for everyone to obey will be rules of infinite complexity telling people to maximise utility in any particular situation in whatever way is appropriate to that situation. (We can see this as an extension of the previous idea, if we count any failure to maximise the good as a disaster to be avoided.)

In order to avoid both of these two objections, Collective Consequentialists seek a middle road between simplistic rules which are subject to the Partial Compliance Objection, and infinitely complex rules which lead to the collapse into Individual Consequentialism. One common response is to appeal to the distinction between "following a rule" and "accepting a rule". The Co-extensionality Objection is said to apply only to the former notion, as there are limits on the complexity of the rules which a community can accept.

The next step is to tease out these constraints on the complexity of a moral code. One such constraint is the following:

The Teachability Constraint : only those rules or dispositions which could be taught to an ordinary human being are allowed. If we are keeping certain institutions fixed, we should allow only those rules or dispositions which our current institutions would be capable of teaching to an ordinary human being. Also, the cost of teaching the rule or disposition, which will presumably increase in proportion to its complexity, must be borne in mind when assessing possible societies especially if we are assuming that the total resources available within society are more or less fixed.

Versions of this constraint have been advocated by (amongst others) Richard Brandt and Brad Hooker.⁶⁵ One problem with this approach is that teachability is a function, *inter alia*,

⁶⁵See Brandt 1992a, 1992b and 1992c, and Hooker 1990 and 1994a.

of time. As Brandt notes with approval, what is teachable in future generations may not be teachable now. The limits of teachability are fluid. Altering the moral code of a society will alter or expand the set of moral codes which could be taught to the next generation. The moral code of a society affects and is affected by the social institutions, behavioural dispositions and educational practices which are current in that society at any one time. The question for those who adopt the teachability constraint is thus "Teachability when?" Are we interested in what is teachable in the first generation, or what is teachable in the second, or ...? There seem to be two basic options: long term teachability and short term teachability, each of which will be dealt with in turn.

2.2.4.1: Teachability in the Long Term

Brandt's own teachability constraint construes the notion of teachability as applying in the long term.⁶⁶ We are interested not in rules which could be taught to human beings now, but in rules which could be taught in the long term in a society which sought to have good maximising rules. For instance, assume that we are trying to maximise the amount of good in our society. We decide that, other things being equal, there will be more good if people are more altruistic. We then set out at each generation to make the next generation more altruistic. Assume also that at each stage future improvements in altruism will only be chosen if these seem to improve the amount of good. Increases in altruism which are too costly to inculcate or produce will not be chosen. We can then measure the degree of altruism in such a society at any point in time. We then say that the optimal degree of altruism, given that we are limiting our attention to rules which could be taught in the long term, will be the degree of altruism which is present in the long term in this society.

We'll now compare the optimal level of altruism to three other degrees of altruism.

⁶⁶See, for instance, Brandt 1992b, p. 125.

- (C) The degree of altruism demanded by Individual Consequentialism of present agents in their actual situations.
- (P) The maximum degree of altruism which is psychologically possible for human beings who have been brought up in our (current) actual society.⁶⁷
- (D) The maximum degree of altruism which it would be reasonable for any moral theory to demand of agents in our situation at present.

We then have three questions for the Collective Consequentialists, a positive answer to any of which will undermine their claim that their theory is not unreasonably demanding.:

- (1) Is the optimal level of altruism greater than the maximum degree of altruism which it would be reasonable for any moral theory to demand?
- (2) Is the optimal level of altruism greater than the maximum degree of altruism which is psychologically possible for (current) human beings?⁶⁸

⁶⁷There are many different interpretations of "P", depending on how we cash out the phrase "psychologically possible".

⁶⁸We should note that the maximum degree of altruism which it would be reasonable for any moral theory to demand and the maximum degree of altruism which is psychologically possible for human beings are unlikely to be identical. In fact, the former should be significantly less than the latter. The demands which we regard as reasonable for morality to make will be less than the limits of our psychology. As I argued in Chapter ONE above, there must be some limit beyond which we are not required to go, even though we could. Without such a limit, there will be no room for supererogation or moral choice, no breathing space in the moral life. The possibility that the optimal level of altruism is greater than the maximum degree of altruism which is psychologically possible for (current) human beings is thus the greater threat to Collective Consequentialism.

(3) Is the optimal level of altruism greater than (or equal to) the degree of altruism demanded by Individual Consequentialism?⁶⁹

We must now ask if the Collective Consequentialist can give adequate (i.e., negative) answers to these questions? Any such answers will naturally be extremely speculative. Even if we assume that there is a limit to the degree of altruism which is teachable in the long term, it would be extremely difficult to locate that limit. In particular, we need to determine which cultural, social, technological, or biological changes will be admitted into our consideration of possible future societies, and which will not. For instance, should we countenance the possibility of a society in which people are made much more altruistic from one generation to the next by means of genetic engineering? If so, then the optimal level of altruism seems very likely to be high. If not, why not?

I do not propose to answer these questions in detail at this point. Instead, I want to make some general comments. The questions the Collective Consequentialist must answer are extremely complex and involved. It seems possible that many of them may not even admit of any precise formulation. However, without reasonably specific answers to these questions, the Collective Consequentialist cannot assure us that she will be able to answer our three principal questions in the negative. Without such assurance, the Collective Consequentialist cannot show that her theory is an acceptable alternative to Individual Consequentialism. Finally, whatever their success in answering them, the very fact that Collective Consequentialists must pose these complicated questions raises serious doubts about their theories. Intricate speculation about the possible future course of biotechnological and socio-cultural evolution simply should not bear so strongly on the

⁶⁹This will most clearly be the case if the optimal level of altruism occurs when everyone treats their own interests exactly on a par with the interests of everyone else. In such a society, everyone will effectively be an Individual Act Consequentialist. As Collective Consequentialism is usually put forward as an alternative, in particular a less demanding alternative, to Individual Consequentialism, the possibility that the optimal level of altruism is no less than the degree of altruism demanded by Individual Consequentialism is a serious threat to Collective Consequentialism.

question of how we, in the actual world at present, ought to respond to pressing moral problems. If Collective Consequentialists are required to answer these questions, then their approach is misguided.⁷⁰

⁷⁰Historically, the notion of long-term teachability may have entered the philosophical literature in the Nineteenth Century via the set of controversies regarding the moral significance of evolutionary theory. We can perhaps best see this by examining the views of Herbert Spencer (which are developed in Spencer 1851, 1862, 1879, 1881 and 1892). Spencer's moral theory is based on a particular view of evolution. His general assumption is that evolution consists of the removal of conflict and a gradual increase in harmony and homogeneity. At the molecular level, this is reflected in the reduction in conflict between individual molecules and the creation of stable, integrated objects. At the social level, the evolution of human society consists in the gradual and inevitable decline in conflict between the interests of one person and another, and the decline of similar conflicts between the interests of different groups. The end result will be a situation in which everyone's interests are in complete harmony. There will thus be no conflict between egoism and altruism, or between either and the demands of morality. What is in one person's interest will be in everyone's interest, and vice versa. Having established the historical inevitability of universal harmony, Spencer then claims that what we ought to do now in our current situation is to obey the moral rules which will be followed in the state of harmony in the future. Spencer's view is extremely optimistic on two counts. First, there is optimism about the result. He assumes that the situation of complete harmony is inevitable. Secondly, there is an optimism of proximity. Not only is this end result inevitable, but it is not particularly far off. The state of complete harmony will be realised once the means of production are sufficient to provide for all the wants of everybody. In the latter part of the Nineteenth Century, Spencer appears to think that such a time is not very far off. This position is open to two basic objections. The first, and perhaps most significant, was presented very forcibly by Sidgwick (see Sidgwick 1886b, 1892 and 1907, p. 470). This is the claim that Spencer's story is simply irrelevant. Even if we accept that he tells us about the future, why should this give us any reason to act in accordance with the moral code which will be prevalent in the ideal society? I believe that this objection is very powerful and that Spencer has no adequate reply to it. The second objection to Spencer's theory is that his optimism is misplaced. Spencer's view of evolution is a very naive adaptationist view in which evolution is seen as progress, as a process by which organisms are perfectly adapted to their environment, and as a teleological activity with a definite and specific end. Progress in evolution is equated with goodness or desirability. Evolution, in Spencer's theory, becomes a surrogate for God. Such a simple, optimistic account would find little favour today. Few but the most enthusiastic socio-biologists would be happy to treat evolution as an inevitable and flawless progress towards perfection, from which we can simply read off moral judgements. (For a contemporary philosophical critique of this extreme view, see Kitcher 1985. See also the mid-Twentieth Century debate between Julian Huxley and C.D. Broad, summarised in Broad 1971d.) Similarly, Spencer's historical and economic optimism seems old-fashioned and misplaced. He assumes that simply having the means to produce enough to satisfy all the material wants of everybody in the world would be sufficient to remove conflicts of interest. This is because he takes wants and needs as given, as related primarily to material goods, and hence as capable of universal and complete satisfaction. The history of the Twentieth Century suggests that a more realistic account of needs and wants would need to make room for the Hobbesian concept of "eminence". What people want is not just to have enough stuff or to have the right stuff, but to have more stuff than other people, and in particular to have scarce things which other people want but can't have. Modern capitalist economies seem to generate their own wants so that some wants always remain unmet. The fact that the means of production are sufficient to meet everyone's basic needs does not guarantee the elimination of conflict. Although Spencer's formulation is scientific, evolutionary and secular, he still retains the very traditional theistic view that the universe is ordered for the purposes and benefits of human beings. The world, to put it simply, is a good place. This quasi religious outlook is the foundation of Spencer's optimism.

There is, of course, a crucial dissimilarity between Spencer's position and that of the Collective Consequentialist. Spencer discusses what will happen in the future and then says that we should model our behaviour on the rules which will become prevalent in that actual society of the future. By contrast, the Collective Consequentialist discusses possible futures in order to help us to draw up the realm of possible societies. Teachability itself does not tell us what we should do, it merely tells us which set of possible worlds we are allowed to choose from. However, the Collective Consequentialist appears to assume that the

2.2.4.2: Teachability in the Short Term

Given the problems which surround a Collective Consequentialism based on teachability in the long term, it seems reasonable to explore the possibility of teachability in the short term. On this view, Collective Consequentialists will limit themselves to the consideration of those rules which could become widely accepted within our society in the near future. A number of questions then arise. The first is: How short is the "short term"? Are we interested in those sets of rules which could be taught to us, or to our children, or to our children's children, or ... ? If we choose to stop at the n -th generation (where n is between 1 and infinity), then we will need some compelling reason for not choosing $n-1$ or $n+1$. The second question is: What do we keep constant, and what do we vary, when we're asking what is teachable? Are we asking what is teachable given the exact institutions we have, given institutions of the general form we have, or given possible institutions which might grow in the next two or three generations out of the institutions we have given the institutional dynamics we have? Are we talking about our own particular psychology, or idealised human psychology, or general human psychology? About our own culture, or any possible culture, or a culture somewhat like our own, or one somewhat like what our culture will or might turn into? If the short term teachability theorist is to have a reasonably definite theory, then these and other questions need to be given relatively definite answers.

A similar question is: Who will do the teaching? Are we looking for rules which we could teach, or for rules which could be taught by someone in our society (teachers,

optimal level of altruism in the long term will not be too much greater than current levels of altruism. This assumption, which seems vital if the Collective Consequentialist is to defeat the Demandingness Objection, mirrors Spencer's optimistic claim that in the ideal society of the future, which is not too far off, people won't be required to make much greater sacrifices than they currently do, as their interests will have converged with those of everyone else. The Collective Consequentialist's optimistic faith in teachability as a middle road between the demands of Individual Consequentialism and the absurdity of rule worship, may thus be a hangover from the more general Nineteenth Century optimism of Spencer and other earlier secular evolutionists.

educationalists, moral philosophers, politicians), or for rules which could only be taught by perfect teachers? Again, relatively definite answers are required if the theory is to have sufficient content to reply to the Demandingness Objection.

On the face of it, it seems as if the short term teachability theorist faces a choice between three basic options:

(1) A theory which is dependent on current institutions, current psychology and current social practices. For instance, we might say that a set of rules only counts as teachable if it is one which we could actually teach to our children. Here the teachability theorist faces a dilemma. On the one hand, a theory which is too closely tied to current institutions and practices will be unable to criticise those institutions or practices. On the other hand, we have no guarantee that a theory which is only loosely based on such institutions will not be significantly more demanding than they are (and, hence, no guarantee that it will not be unreasonably demanding).

(2) A theory which is dependent on some specific idealisation, some particular utopian vision of what is teachable. The problem then would be to justify this specific idealisation in all its complexity, as opposed to every other specific idealisation. Once again, we have no guarantee that such a theory won't be unreasonably demanding.

(3) A theory which is vague, indeterminate, or impossibly difficult to cash-out. For instance, if we limit ourselves to those sets of rules which human beings are psychologically capable of learning, then, even if there is a definite answer to the question of where those limits lie, it seems very unlikely, given our reluctance or inability to perform the relevant and necessary experiments, that we would ever determine that answer with sufficient accuracy to know whether or not our theory was too demanding.

This is not of course to say that short term teachability cannot be made to work, merely that there is no particular reason to believe that it can. Unfortunately, most Collective Consequentialists merely sketch their accounts of short term teachability. The questions and difficulties I raise are more often noted or glossed over than addressed. Nowhere, to my knowledge, are we given a detailed account of how such a theory would be worked through in practice. Of course, a perfectly detailed account would be impossible to construct. Some degree of abstract hand-waving is inevitable. However, without a good deal more detail than has as yet been forthcoming, we have no particular reason to believe that a viable account of short term teachability is even a coherent possibility. It is thus reasonable to be sceptical of the weight which some theorists seek to place on the notion of teachability.

Both options for the teachability theory are extremely problematic. Neither long term teachability nor short term teachability offers a credible basis for a Collective Consequentialism. Faced with such difficulties, Collective Consequentialists might be tempted to replace teachability with another constraint on the complexity of moral rules. Initially, there seems no reason why another such constraint might not be available. However, if we accept the plausible view that the notion of a rule, particularly the notion of a moral rule, is necessarily tied up to its being capable of being learnt by human beings, then it is not so clear that we could have a constraint on the complexity of such rules which was completely independent of the notion of teachability. Any constraint must appeal to psychological features of human beings, not merely to formal features of rules. Yet, any such psychological constraint will face the same problems as the teachability constraint. It is likely to be either too closely tied to existing behaviour to permit it to criticise that behaviour, or too demanding to be acceptable, or too vague to be of any use.

2.2.5: Why Collective Consequentialism isn't really fair

Collective Consequentialists often defend their theory on the grounds of "fairness". In giving 10%, I do my share. My giving another 5% would only improve aggregate welfare because other people aren't doing their share. Why should I be punished (by being made to donate an extra 5%) just because other people are avoiding their moral responsibilities? That would be unfair. It's only reasonable to require me to do my share. Otherwise, other people can exploit me by not doing theirs.

This familiar argument is based on a false analogy. The notion of "doing my fair share" is most naturally applied to the provision of public goods, where everyone benefits, regardless of whether or not they have contributed. If I do my share of producing the good, others who fail to do their share can be said to "free ride" on my provision. By refusing to do their share of the work, they force me to do more than my share. This is unfair.

Famine relief is clearly not a standard public good. There are two crucial differences. The first is that, while the "others" who refuse to do their fair share do benefit from not making donations (by having more money left over to spend on themselves), they do not benefit from my doing more than my fair share. They are not free-riding on my extra work. A related difference is that, in the case of famine relief, the "others" who will suffer if I only do my "fair" share, rather than giving as much as I can, are innocent third parties: namely, the famine victims. It would be absurd to accuse *them* of free-riding if I donate 15% rather than 10%.

In the famine case, then, the notion of "fair share", if it is to be applied at all, should have a different sense than in the public good case. I suggest that, with famine relief, my fair share should be dependent upon two factors: the relative disadvantage to my welfare, interests or

projects of parting with various amounts of wealth, and the benefits which various donations would provide for the famine victims. I think we are more inclined, in cases of this kind, to say that someone has given their fair share if they have, say, “given ‘til it hurts”, than if they have given less than they could have simply on the grounds that, if everyone had given the same, the problem would have been solved. Collective Consequentialists are right to emphasise the importance of “doing one’s fair share”, but wrong to see this as an argument in favour of Collective Consequentialism.⁷¹

2.2.6: What's Wrong with Homogeneity

The Homogeneity Thesis holds that the best society would be one in which everyone has the same moral character.⁷² This thesis might be false. As a simplistic example, consider the following story:

Dictator's Education: Dictator, a benevolent ruler, is responsible for education in a large nation state. Dictator is a Consequentialist, so she wishes her education system to maximise the amount of good produced. In general, Dictator thinks it would be best if people were kind, well-meaning non-Consequentialists. However, there are two small classes of people who should be different: politicians and soldiers. Politicians will make generalised decisions about how resources are to be allocated within society. Dictator reasons that these people should make their decisions solely on the basis of impersonal cost-benefit analyses. Soldiers are necessary to protect the society from hostile invasion. Dictator believes that violent, specialised mercenaries make the best soldiers. Dictator thus wants three types of moral character: citizen, politician and soldier. The total population available in Dictator's

⁷¹Murphy's account of benevolence may rest on similar confusions. See Chapter THREE. Perhaps the intuition about fairness is better explained by appeal to the moral significance of the notion of cost to the agent, which is discussed at considerable length in Chapters FOUR, FIVE and SEVEN below.

⁷²See pp. 38-39 above.

society in each generation is P. Dictator decides that, when the next generation graduates, society will need X politicians and S soldiers. Three schools are established: Calculator College (with X pupils), Warrior College (with S pupils) and Ordinary College (with $P - X - S$ pupils). Each school produces a different moral outlook. Graduates of Calculator College are hard-nosed Act Utilitarians, making decisions on the basis of impersonal cost-benefit analyses. Graduates of Warrior College are vicious, highly-disciplined, order-obeying killers. Graduates of Ordinary College believe that morality is largely a matter of personal, one-to-one relationships with people one knows individually. Calculators are thus ideally suited to the politician role in Dictator's society, Warriors are fitted for the soldier role, while Ordinaries are the sort of people you generally want in your society. By creating two groups with radically different moral views, Dictator creates a better system (from a Consequentialist perspective) than would have been possible if everyone had the same moral character.

Defenders of the Homogeneity Thesis must argue that Dictator is wrong. There is a single moral character which, if universally adopted, would be preferable to (or, at least, no worse than) Dictator's non-homogenous solution. Likely options include the following:

Universal Calculators : everyone attends Calculator College. Everyone is an Act Utilitarian.

Universal Warriors : everyone attends Warrior College. Everyone is a trained and obedient killer.

Universal Ordinary : Everyone attends Ordinary College. Everyone gives moral priority to those in their immediate proximity, or to those who stand in a special relation to them.

Universal Flexibles : a Flexible is a versatile moral character. She is able to modify her behaviour according to the social role in which she finds herself cast. In particular, Flexibles will generally be disposed to perform cost-benefit analyses before acting. So they will be well suited to the role of politician. On the other hand, Flexibles believe that cost-benefit analysis is inappropriate in ordinary life. In their role as ordinary citizens, Flexibles will thus behave exactly like Ordinaries. Finally, Flexibles believe that (some) ordinary standards of decency are inappropriate in a state of war. So they are capable of behaving like Warriors if the need arises.

In effect, a population of Flexibles will behave exactly like Dictator's own population, with the additional advantage of improved flexibility, as Flexibles can switch roles in a way that Calculators, Warriors and Ordinaries cannot. If a society consisting exclusively of Flexible characters is feasible, it must be preferable to Dictator's non-homogenous combination.

In other aspects of their theory, Collective Consequentialists are very concerned with the limitations and particularities of the human condition. Their principal objection to Individual Consequentialism is usually that it ignores these real-world features. Collective Consequentialists don't ask which character would produce the best results if everyone had it in the best society in which it would be logically possible for rational agents to live . Rather, they ask which character would produce the best results in the type of society in which contemporary human beings live. In order to remain true to this tradition of psychological realism, proponents of the Homogeneity Thesis must argue that Flexible is a psychologically possible character for a human being to possess.

Our discussion of Dictator's Education shows that the issues surrounding the Homogeneity Thesis are very complicated, even in this simplified situation. Introducing the complexity of the real world would muddy the waters still further. I am not at all sure what the combined

effect of all these complications would be. At this stage, I suggest that, at the very least, we cannot be confident that introducing real life complexity would necessarily vindicate the Homogeneity Thesis.

Let us assume that the Homogeneity Thesis is false. That is, the possible society which contains the most value is one in which not everyone has the same moral character, or follows the same moral rules. Suppose that, by some miracle, we lived in that world (call it W). In W, everyone's moral character is such that there is no re-allocation of moral characters which would produce a better outcome. As the Homogeneity Thesis is false, different inhabitants of W will have different moral characters and incompatible moral beliefs. Let us focus on Perfect, an inhabitant of W who has character C+. Let us now construct the set of all possible societies in which everyone has the same moral character. Out of this set, we select the world which contains the most value (call this world HomW). All the inhabitants of HomW have the same character (call it C). Ex hypothesi, W contains more value than any world in which Perfect has any character other than C+. A fortiori, the Consequentialist value of W is greater than that of HomW. So, if Perfect had character C, her world would be a worse place. Even if everyone else had character C as well, W would still be a worse place than it is.

One day, Perfect finds herself in a situation where she has two choices: the c-act (what a person with character C would do) and the c+-act (what a person with character C+ would do). What is the right thing for Perfect to do? Proponents of the Homogeneity Thesis must argue that Perfect should do the c-act. Perfect should act as if she possessed a certain moral character, even though her actually possessing that moral character would have made the overall outcome worse in moral terms, regardless of how others behaved. This claim is by no means absurd. Many non-Consequentialists hold similar views. But what surely is

absurd is to base this claim on an argument from the moral value of outcomes. This view is not one which Consequentialists can hold.

It seems that Collective Consequentialists must accept that Perfect should do the c+-act rather than the c-act. However, this would introduce an unwelcome complexity into their theory. If we are to emulate the behaviour of the inhabitants of Utopia, and if we accept that different Utopians will behave differently, then we must ask which Utopian(s) we are to emulate. The choice of A's moral role model (which we first raised in section 2.1.3 above) is no longer trivial. As Collective Consequentialism already contains more than its fair share of indeterminacies, the introduction of another dimension of complexity is probably to be avoided!

The most effective response for Collective Consequentialists to make out of this point would thus probably be to defend the Homogeneity Thesis. One way to do this would be to appeal to a distinction between the following three interpretations of that thesis:

The Same Beliefs Interpretation : Everyone in Utopia has the same moral beliefs.

The Same Rules Interpretation : Everyone in Utopia accepts the same moral rules.

The Same Character Interpretation : Everyone in Utopia has the same moral character.

My discussion has focused on the Same Character Interpretation. Defenders of the Homogeneity Thesis might argue that I have attacked a straw position. While the Same Character Interpretation has been defeated, the Same Beliefs and Same Rules Interpretations remain intact, and these are more relevant to the assessment of Collective Consequentialism. They would then claim that these other interpretations survive the attack

on the Same Character Interpretation. For instance, we could retell the Dictator's Education story so that, although different people had different moral characters, everyone shared the same moral beliefs. Similarly, if we cast our moral rules as hypothetical statements concerning how individuals ought to behave in certain tightly-defined situations, then we can tell a story in which a society of people with different moral characters all accept, and perhaps even follow, exactly the same moral rules.

While I admit that the Same Beliefs and Same Rules Interpretations do seem more plausible than the Same Character Interpretation, I have two principal objections to this response. In the first place, recall that, under Collective Consequentialism, we are primarily interested in discovering what an inhabitant of Utopia would do in certain circumstances. We are not just interested in moral beliefs. Yet, if we are told that, although Utopians have the same moral beliefs, they have different moral characters, then we must assume that in some situations different Utopians would act differently. Otherwise, it is not clear what is meant by admitting that they had different moral characters. The indeterminacy problems which arise under the Same Character Interpretation of the Homogeneity Thesis would thus recur under the Same Beliefs Interpretation. It is the presence of indeterminacy and not the Homogeneity Thesis per se which is the real problem for Collective Consequentialism.

My second reservation is related to the first. Even under the Same Beliefs Interpretation, the Homogeneity Thesis is by no means obvious. For the moral beliefs of Utopians to be of any relevance to us, those beliefs must be linked to the behaviour of Utopians. But then, how can we be sure that Utopia won't be a place where different people have different beliefs? Assume that the behaviour described in Dictator's Education is the optimum pattern of behaviour. Utopia will thus be the closest possible world in which this pattern of behaviour is present. In other words, our description of Utopia will consist of the most plausible story we can tell about why the people of Dictator's Education behave as they do.

The conceptual links between behaviour, moral character and moral beliefs place limits on the range of explanatory stories which can be told. In particular, many possible stories in which people with identical moral beliefs have significantly different moral characters will be ruled out as psychologically implausible. It is then very possible that the best explanation of the behaviour in Dictator's Education will be one in which different people have different moral beliefs. Yet, if this is so, then the falsity of the Same Character Interpretation implies the falsity of the Same Beliefs Interpretation. The Homogeneity Thesis would then be defeated under any interpretation.

To remove my reservations, defenders of the Homogeneity Thesis would need to tell a plausible story about a society in which, although moral beliefs were strongly connected to behaviour, homogeneous moral beliefs co-existed with very different moral characters. I doubt that such a story could be told.

2.2.7: The Problem of Alien Situations

The central question of Collective Consequentialism is the following: What would an inhabitant of Utopia do in my situation? We have seen that this question is very difficult to answer. I wish to explore another difficulty: the possibility that, from the point of view of a Utopian, my situation is an impossible one.

My point is not merely that Utopians might view our situation as unfortunate or undesirable. The real problem is that, when we ask how they would act in a situation such as Affluent's, we are asking how they would act in a situation which people with their background and dispositions would never encounter, and for which they would be totally unprepared. Even if we could answer this question, why should that answer be of any particular interest to us? What guidance can the (hypothetical) response of an ideal agent to

a situation which is completely alien to their experience provide for those of us for whom that situation is an everyday reality? For instance, Utopians might be particularly sensitive to the sufferings to others. This sensitivity would be very useful in their natural environment, where such sufferings are minimised. However, in a real-life situation, such a disposition might produce paralysing despair. While some degree of remorse or regret may well be appropriate when confronted with moral dilemmas, despair does not seem to be morally required.

This point is not confined to the case of famine relief. Many (if not most) of our difficult moral choices concern situations which would be totally foreign to a Utopian: situations arising from the moral or organisational imperfections of individuals (sometimes ourselves) and institutions. It's not clear that the Collective Consequentialist's strategy can offer any reliable guidance on these choices at all.⁷³

Related dilemmas can arise within a given society, not just with regard to people in foreign lands. Inhabitants of developed countries with underdeveloped welfare states are regularly confronted with people living in poverty. Giving aid to such people whenever one encountered them would crowd-out all other personal projects. However, Utopians are unlikely to encounter beggars, so they may be disposed to help them as much as they can. In Utopia, such a disposition has no negative effects, but following it in the real world would be very demanding.

These points are particularly relevant in light of Hooker's insistence on complexity constraints.⁷⁴ As he rightly points out, there are limits on the complexity of a learnable set of moral rules. The moral rules accepted in a given society will thus be tailored to the

⁷³See also section 2.2.2 above.

⁷⁴See also section 2.2.4 above.

circumstances of that society. Rules which deal with common situations will be very detailed, whereas those which deal with unfamiliar events will be comparatively unsophisticated. In Utopia, poverty is rare (if not entirely absent). Accordingly, Utopians will have very crude moral rules for dealing with poverty. It seems unlikely that these rules will be the most appropriate ones for us to adopt in our own society.

2.3: Conclusions

I conclude both that Collective Consequentialism cannot provide an adequate response to the Demandingness Objection and that Collective Consequentialism is not an acceptable moral theory. My objections to Collective Consequentialism also raise doubts about any approach based on the Utopian model, whether Consequentialist or not.

Chapter THREE: Liam Murphy's Cooperative Principle

In this chapter, I discuss a recent alternative to the Collective Consequentialism discussed in Chapter TWO.

Liam Murphy suggests the following condition on principles of beneficence:⁷⁵

Compliance Condition : A principle of beneficence should not increase its demands on agents as expected compliance with the principle by other agents decreases.

Murphy then seeks a principle which meets this constraint. He suggests the following:

Cooperative Principle of Beneficence : Each agent is required to act optimally - to perform the action that makes the outcome best - except in situations of partial compliance with this principle. In situations of partial compliance it is permissible to act optimally, but the sacrifice each agent is required to make is limited to the level of sacrifice that would be optimal if the situation were one of full compliance. Of the actions which require no more than this level of sacrifice, an agent is permitted to perform only that action which makes the outcome best. The agent is then required to perform either that action, or some other action which makes the outcome just as good.⁷⁶

I shall argue that this principle is open to a number objections. I shall suggest that any other principle which satisfies Murphy's Compliance Condition will face similar objections.

⁷⁵See Murphy 1993.

⁷⁶I have slightly reworded Murphy's published formulation, but I think my version captures his intentions.

Murphy's position has obvious similarities with Collective Consequentialism. In terms of our earlier framework, we can split Murphy's theory into two halves:

(1) The definition of Utopia.

(2) Specify a function whose domain is a set of ordered pairs of agents and possible situations, and whose range is the set of acts available to the agent in situation S. The value of this function will then be the right thing for our agent (A) to do in S.

Murphy's account of step (1) is analogous Hooker's. Utopia is an idealised world of full compliance, where everyone acts optimally. The principal difference lies at step (2).

Murphy uses Utopia, not to determine what A should do in S, but rather to determine how much sacrifice she should bear. This enables Murphy to avoid those of our objections to Collective Consequentialism which were based on the odd results it gave in situations of partial compliance.⁷⁷

However, Murphy's principle is subject to a number of the same objections which defeated Collective Consequentialism, particularly the Wrong Facts Objection (see section 2.2.1 of Chapter TWO) and the Exception Clause Objection (see section 2.2.2 of Chapter TWO).

⁷⁷ Another difference we should note is that, unlike Hooker, Murphy does not appeal to any distinction between following a rule and accepting a rule. Murphy's Utopia is simply a place where everyone follows the rule of acting optimally. This saves Murphy from those of our previous objections which attacked that distinction. However, it also deprives him of those of Hooker's replies which utilised the distinction.

3.1: The Wrong Facts Objection

Recall Affluent's Three Ignorances:⁷⁸

Affluent's First Ignorance : Affluent does not know how many people are starving. All she knows is that she is in one of three situations: Few Poor, Medium Poor, or Many Poor.

Affluent's Second Ignorance : Affluent is drastically misinformed about the size of the developed world. She doesn't know whether she's in Few Rich, Medium Rich or Many Rich.

Affluent's Third Ignorance : Affluent is ignorant of the facts of political economy. She doesn't know if she's living in Developmentland, Senland or Malthusia.

It is clear that, for Murphy, as for the Collective Consequentialist, Affluent's obligations depend upon the precise nature of the optimal global response to famine. Murphy seems doomed to give implausible answers in Affluent's three cases of ignorance.

Murphy is committed to the view that the sacrifice required of an individual is equal to the total good to be done in any particular situation divided by the number of agents available to produce that good. Let us assume, as we did in Chapter TWO, that, in the actual world, full compliance with a rule requiring a donation of 10% of one's income would produce the best possible results. In the actual world, then, Affluent should follow this rule and donate 10%. However, if this is the appropriate individual sacrifice, that must be because the total sacrifice required to produce the required good (namely, the alleviation of poverty and starvation), divided by the number of people who are in a position to make such donations,

⁷⁸See section 2.2.1 of Chapter TWO above.

would equal 10% of the income of the latter. Changing the number of people facing starvation by a factor of fifty would thus lead to a proportionate change in the amount of good to be done, and hence to a fiftyfold increase in the sacrifice required under full compliance. Similarly, decreasing the number of starving people by a factor of fifty will lead to a proportionate decrease in the required sacrifice. Any alteration to the number of affluent people able to collectively alleviate poverty will also lead to a proportionate increase or decrease in the individual sacrifice required. Finally, any change in the efficiency of aid agencies will alter the total sacrifice required to produce any given amount of good, thus leading to a proportionate change in the sacrifice required of each individual. Murphy is thus doomed to provide the same recommendations to Affluent in each of her three ignorances as Hooker did.

Murphy discusses a very similar objection in his paper:⁷⁹

A way of testing this is to think of the case of an extremely well-off minority in an extremely badly off country, isolated from the rest of the world. ... the Cooperative Principle would impose great sacrifice on this minority even under full compliance.

Murphy responds to this objection as follows:

But would we feel that the extreme demands in these cases were objectionable, in just the same way that the demands made on us every day by the Simple Principle [which corresponds to Individual Consequentialism] seem objectionable?

In my objections to Collective Consequentialism in Chapter TWO, I argued that we *would* regard the extreme demands of the Cooperative Principle as objectionable. Indeed, I argued

⁷⁹ibid., p. 289.

there that those demands are, in a sense, more objectionable than the extreme demands of Individual Consequentialism. The latter makes extreme demands of Affluent whatever her circumstances. By contrast, the Cooperative Principle makes extreme demands in some cases but not in others. It is this variation to which the Wrong Facts Objection objects.

We should also note that this tentative response of Murphy's seems far better suited to one side of the Wrong Facts Objection than to the other. Murphy mentions the over-demanding cases, in which fluctuating numbers lead the Cooperative Principle to make extreme demands. This leaves out the under-demanding cases (Few Poor and Many Rich), in which fluctuating numbers lead the Cooperative Principle to make very light demands. Surely we do find it objectionable that a principle requires so little of agents, even when there is still so much to be done.

In combating the under-demanding case, Murphy is on much weaker ground than Collective Consequentialists such as Hooker. This is because the latter focus on rules, whereas Murphy deals merely with levels of sacrifice. In sophisticated forms of Collective Consequentialism, the flexibility of rules is made to do a lot of work.⁸⁰ For instance, Collective Consequentialists may respond to the Wrong Facts Objection by claiming that a rule which would demand little in cases of full compliance may demand quite a lot in situations of partial compliance. Such moves are unavailable to Murphy. Under his principles, if the sacrifice under full compliance would only be one dollar, then no acceptable moral theory can ask Affluent to donate more than one dollar.

The Wrong Facts Objection thus seems even more decisive against Murphy's view than against conventional Collective Consequentialism.

⁸⁰In Chapter TWO, I argued at length that, in fact, too much work is required of this flexibility.

3.2: Natural versus Artificial Disasters

Recall the following four stories:⁸¹

Famine Just Happens : through no action or inaction or fault of any particular person or group of people, some people face starvation or famine.

The Brigands' Famine : the soldiers of one side in a civil war burn all the crops in the territory of the other side.

The Hoarders' Famine : famine arises because those who own the grain refuse to sell it or give it to those who require it.

The Ignorers' Famine : famine results (in part) from the collective inaction of developed countries.

Murphy says that our obligations depend upon what would happen under full compliance with his principle. In *Famine Just Happens*, full compliance would consist of a world-wide disaster-relief program. The sacrifice required of us will thus be equal to that implied by our share in that ideal relief-effort. In each of the other three famines, full compliance would eliminate the famine before it began, as it has only arisen as a result of peoples' failure to act optimally. Murphy's principle would thus require no sacrifice from us at all.

The demands of Murphy's principle thus vary enormously depending upon whether or not the disaster which confronts us is the result of natural forces or human actions. This seems unreasonable. When faced with a disaster, surely my first questions should be "How can I

⁸¹See section 2.2.2 of Chapter TWO.

help?" and (perhaps) "How much will it cost me?" not "Is this the result of an act of God?" (followed by "If not, I'm not obliged to help at all.")

In a footnote to his paper, Murphy seeks to avoid a similar conclusion by suggesting that:⁸²

the Cooperative Principle is accompanied by deontological constraints against violence, thieving, and the like. For in that case we would not characterise the actions of wrongdoers like those in the former Yugoslavia and in Somalia as failures in beneficence.

This move might remove the anomaly of the Brigands' Famine, but it cannot help us in either the Hoarders' Famine or the Ignorers' Famine, unless we include deontological constraints against failing to give discounts to starving customers or failing to make adequate donations to Oxfam. The introduction of such constraints would crowd-out the original benevolence principle. If our deontological constraints rule out all human behaviour which causes (either actively or through omission) human need, then there will be no room left for benevolence. Besides, no moral theory which included such wide-ranging constraints could possibly hope to defeat the Demandingness Objection.

3.3: Is Murphy's theory really undemanding?

If the Cooperative Principle is to be of any use to us, then it must not make very great demands in the actual world. I shall now argue that in fact it does make undue demands.

⁸²ibid., pp. 280-281, footnote 27.

Recall that Hooker claimed that Collective Consequentialism is not too demanding in the actual world. In Section 2.2.3 of Chapter TWO, I rejected this claim. My principal argument was that widespread diminishing marginal returns implied that, while universal donation of 10% from the well-off might be sufficient to eliminate starvation, greater donations would provide further increases in the welfare of the worse-off. The optimal donation level would occur when the benefits to the worse-off balanced the sacrifices of the well-off. To require such a level of donation would be very demanding in the actual world.

Now, Murphy's notion of the maximum sacrifice required under full compliance is similar to Hooker's notion of the optimal donation required under full compliance. In particular, just like the latter, the former is threatened by diminishing marginal returns. A sacrifice equivalent to a donation of 10% may be sufficient (under full compliance) to eliminate starvation, but a much greater level of sacrifice would produce even greater increases in overall welfare. The maximum level of sacrifice will be reached only when benefits to the worse-off are outweighed by the cost to the well-off. This may not occur at precisely the same point as in Hooker's Collective Consequentialism, but the resulting account is still likely to be very demanding.

Murphy's Cooperative Principle will probably not be as demanding in the actual world as Hooker's Collective Consequentialism. As we saw earlier, Murphy's levels of sacrifice are less flexible than Hooker's optimal rule. In other words, while Collective Consequentialism may well demand a higher level of sacrifice under partial compliance than under full compliance, Murphy's theory, by definition, always demands the same sacrifice under partial compliance as under full compliance. On the other hand, this very lack of flexibility leads Murphy's theory to fare even worse than Hooker's against the Wrong Facts Objection. Furthermore, even its increased rigidity is very unlikely to prevent Murphy's Cooperative Principle from being unacceptably demanding in the actual world.

3.4: Does Murphy permit harming for personal gain?

In Chapter FIVE below, I discuss Shelly Kagan's objection that Scheffler's Hybrid View, lacking a distinction between doing and allowing, must permit agents to harm others in pursuit of their personal interests. It looks as if Murphy will be open to an analogous objection, as his theory fixes the maximum level of sacrifice required of all agents without any reference to the details of how that sacrifice comes about. For instance, if the maximum sacrifice required of me is equivalent to a loss of \$1,000, then, under Murphy's Cooperative Principle, I am allowed to perform an action which involves no greater sacrifice than that. Suppose that, for some reason, the only way I can avoid a loss much greater than \$1,000 is to kill my ageing uncle and inherit his fortune. (Perhaps my own fortune is about to disappear in a market slump.) I am then allowed (though, on my formulation of Murphy's principle, I am not required) to kill my uncle for personal gain.⁸³

I discuss Kagan's argument at some length in Chapter FIVE, along with a number of possible replies available to Scheffler. At this stage I shall say only that if Kagan's objection succeeds against Scheffler's Hybrid View, then it will also succeed against Murphy's Cooperative Principle. Similarly, most of the replies open to Scheffler would also be open to Murphy.

Murphy's own reply to this type of objection would probably be that his Cooperative Principle is designed not to be a whole moral theory but merely to be one principle among many. If we wish to make room for a genuine distinction between doing and allowing, then

⁸³It may seem that Hooker's Collective Consequentialism will be subject to similar objections. However, the greater flexibility of Collective Consequentialism comes to its rescue here, as the Collective Consequentialist can argue that the Utopian moral code will include a rule against murdering close relatives for personal financial gain.

we should simply add suitable no-harm-for-personal-gain principles to our theory. This would be enough to dissolve Kagan's objection. For our purposes, however, such a move represents too great a departure from ordinary Consequentialism, as we seek replies to the Demandingness Objection which do without explicit invocation of the distinction between doing and allowing. Once we introduce such principles we might also find that the Cooperative Principle itself became redundant as a mitigater of demandingness, as all the work will be done by the additional principles.

3.5: Murphy's Conception of Benevolence

Murphy's most natural response to our first two objections will be that they are not objections at all. Given his conception of benevolence, its demands should vary with the numbers in Affluent's Ignorances, and with the sources of famine in our four famine tales.

Murphy's explains his conception as follows:⁸⁴

Beneficence can be understood in terms of a shared cooperative aim ... if we both have a cooperative aim to promote the good ... we do not see ourselves as engaged in separate solitary enterprises. ... Each of us does not, strictly speaking, aim to promote the good. Each sees himself as *working with others* to promote the good. Thus, the best way to describe the aim of each might be: 'to promote the good together with others'.

Such a conception of benevolence is certainly appealing in public good cases. As Murphy notes:⁸⁵

⁸⁴ibid., pp. 285-286.

⁸⁵ibid., p. 288.

insofar as beneficence is ... a mutually beneficial project, it is natural to resist taking on the shares of people who could contribute to the project but do not.

However, as I pointed out in section 2.2.5 of Chapter TWO (and as Murphy admits), many situations of benevolence are not mutually advantageous. In such cases, Murphy's conception seems, on the face of it, rather odd. My aim in donating to Oxfam is to promote the good (or, more accurately, to relieve some particular suffering). Certainly, I would like all my fellow citizens to cooperate with me in this aim, but only because such cooperation would further my own aim.⁸⁶

Murphy's conception of benevolence has strong analogies with the claim that Collective Consequentialism is fair. I argued in section 2.2.5 of Chapter TWO that the latter claim is only plausible if we restrict our attention to public good cases. We have just seen that the same is true of Murphy's conception. Murphy is thus best interpreted as providing not a genuine account of benevolence, but rather a specific account of benevolence in a certain class of situations. Unfortunately, the Demandingness Objection arises most starkly in the very situations to which Murphy's conception seems least well suited, such as situations in which many distant people are starving. An account of benevolence which is not applicable to those cases will be unable to provide a solution to the Demandingness Objection.

Murphy would probably deny that his account of benevolence applies only to a limited range of cases. He certainly presents his account as applying more generally. Perhaps we

⁸⁶Also, their cooperation might further other aims I have, such as the fostering of community spirit. However, such aims are clearly secondary. We might note, in passing, that many actions performed for "moral" reasons also proceed from non-moral motives. For instance, my desire to promote the good may be combined with a non-moral desire to be part of a group which shares that aim. If no such group is available, then the sacrifice involved in making a given contribution to the good will be far greater, as I will not be compensated with the benefits of belonging to a group of like-minded do-gooders. In such a situation, I might well be required to make a greater contribution in a cooperative setting than in a non-cooperative one, because the overall level of sacrifice involved is the same.

are faced here with a difference in intuitive responses. Murphy finds it appealing to regard benevolence as an essentially cooperative enterprise across the board, whereas I only find this plausible in cases of mutual advantage. I have no knock-down argument to present to those who, after reading this chapter, still share Murphy's intuition. However, anyone who does not share that intuition, or even anyone who has doubts about it, will want to pursue the search for an alternative solution to the Demandingness Objection.

Chapter FOUR: Slote's Satisficing Consequentialism

In this chapter, I discuss a prominent moral theory which responds to the Demandingness Objection by incorporating an explicit departure from Consequentialism within a broadly Consequentialist framework, namely Michael Slote's Satisficing Consequentialism.⁸⁷

4.1: Slote's Argument

Michael Slote defends a theory called "Satisficing Consequentialism".⁸⁸ This is based on the economic notion of satisficing:

(rational) economic agents may sometimes seek less than the best, may sometimes choose what is good enough (in terms of the agent's own interests), without regard for whether what they have chosen is the best thing (outcome) available in the circumstances.

Call this behaviour "economic satisficing". Similarly, moral satisficers claim that:

moral agents may sometimes seek less than the best, may sometimes choose what is good enough (in terms of moral value), without regard for whether what they have chosen is the best thing (outcome) available in the circumstances.

The most familiar form of moral satisficing is:

⁸⁷An earlier version of this chapter appeared as Mulgan 1993a.

⁸⁸In Slote 1985 and Slote 1991.

Strategic Moral Satisficing : maximising agents should adopt satisficing procedures because these will produce better results than maximising procedures. For instance, if time is short, it may be best to just take the first "good enough" option that turns up, rather than laboriously calculating the exact value of each option.

Many Consequentialists accept some form of Strategic Moral Satisficing. Slote, however, also accepts the following:

Blatant Moral Satisficing : even when the agent already knows which particular act is the best (from a Consequentialist perspective), she is still perfectly justified in selecting any other act which is good enough.

Slote thus supports satisficing even when it is not a strategy for maximising. In other words, Slote advocates satisficing as a criterion of the moral rightness of acts, not just as a decision-procedure which might be recommended by those who adopt a maximising criterion of rightness.

Slote begins with the general suggestion that Consequentialist morality should be analogous to economic rationality. This leads him to the following two theses:⁸⁹

The Analogy Thesis : If blatant economic satisficing is a rational way for economic agents to pursue their goals in certain circumstances, then Consequentialism should accept blatant moral satisficing in similar circumstances.

The Economic Thesis : blatant economic satisficing is often a rational way for economic agents to act.

⁸⁹The formulations and names of all theses and principles are my own.

These theses combine to produce the following further claim:

The Satisficing Thesis : Consequentialists must accept some blatant moral satisficing.

This thesis in turn allows Slote to reply to the Demandingness Objection. Slote would agree that the objection defeats Maximising Consequentialism. However, he claims that Satisficing Consequentialists can defeat the objection, by establishing the following thesis:

The Undemanding Thesis : By employing an appropriate notion of the "morally good enough", Consequentialists can use blatant moral satisficing to defeat the Demandingness Objection.

In what follows I shall focus on this Undemanding Thesis.

4.2: Common Objections to Slote

One common response to Slote's argument is to reject the Economic Thesis, to deny that blatant economic satisficing is ever rational. Most economists advocate only Strategic Satisficing, and not Blatant Satisficing. For instance, in his discussion of rational institutional behaviour, Herbert Simon concludes that it's not always rational to seek to maximise and that institutions should instead be satisfied with attaining a certain "aspiration level".⁹⁰ However, he explicitly says that "when a firm has alternatives open to it that are at or above its aspiration level, it will choose the best of those known to be available."

⁹⁰See Simon 1976.

Simon is thus rejecting Blatant Satisficing. Following Simon, Philip Pettit suggests that Slote's position is incoherent, as "it is not clear what it can mean to rank A above B if when other things are equal one insists on choosing B".⁹¹ Pettit supports this attack as follows:

Suppose that an agent ranks A above B ... whereas he could have given a reason for choosing A ... he can give no reason for choosing B. It will not do for him to say simply that B is good enough. That might be a reason for him to go for B, if he were unaware of the nature or value of alternatives. It is not a reason for him to choose B rather than A.⁹²

In response, Slote argues that we *can* find reasons for the behaviour of the blatant economic satisficer.⁹³ He offers two candidate reasons:

- (1) The agent doesn't need more; and
- (2) The agent is perfectly satisfied with her present state.

⁹¹Pettit 1984, at p. 173. At this point, we should note an ambiguity in the notion of Satisficing. This is the distinction between the strong claim that an agent can have a reason to choose B over A, when A is better than B, and the weaker claim that, although the agent can give no reason for choosing B over A, she is still (rationally) permitted to pick B instead of A. As originally formulated, Pettit's argument only tells against the stronger claim, as proponents of the weaker claim will accept that the agent has no reason to choose B rather than A. However, Slote often speaks as if he means to make the stronger claim, especially as some of his examples appear to turn on the idea that moderation is a virtue (and, hence, a source of reasons for action). More importantly, Pettit's argument can be modified to deal with the weaker claim as well. If we accept that A's being better than B gives the agent a reason for choosing A over B, then proponents of the weaker view must give us some reason why, in the absence of a counter-reason for choosing B over A, the agent is allowed to pick B. Pettit's argument would then be that no such reason can be given, as the idea of picking A over B without reason only makes sense if A and B are effectively identical (as, for instance, when one picks one of two identical cans of soup). As the same reasons are likely to be suggested in either case, it will be convenient in the text to retain Pettit's original formulation.

⁹²ibid., p. 172.

⁹³These are given in Slote 1991, p. 21.

Slote argues that either of these could provide an agent with a reason to choose B rather than A, even when the only salient difference between the two options is that A is "better" (for the agent) than B.

I propose not to discuss Slote's reasons at this point. Rather, I shall use them later on to impale him on the horns of a dilemma. I shall contend that, if Slote uses his two reasons to defend the claim that blatant economic satisficing is sometimes rational, then he can no longer claim that Consequentialist morality is analogous to economic rationality. This is because, whatever the fate of Slote's reasons in their original context, they have no acceptable moral analogues. Slote will be forced to reject either the Analogy Thesis or the Economic Thesis. Either way, his arguments for the Undemanding Thesis will be dissolved.

4.3: A Moral Counter-example

Slote's claim is that blatant moral satisficing can be used to rebut the Demandingness Objection. This objection arises because there are some situations in which agents can leave room for their own projects only by producing significantly less good than they could. Maximising Consequentialism cannot permit this. Slote replies that Satisficing Consequentialism leaves room for personal interests and projects. He must establish that it is sometimes acceptable to do much less good than one could, simply on the grounds that one has produced enough good already. I shall argue that we cannot accept this. Consider the following story:

The Magic Game : Achilles is locked in a room, with a single door. In front of him is a computer screen, with a number on it (call it n), and a numerical keypad. Achilles knows that n is the number of people who are living below the poverty line. He also knows that,

as soon as he enters a number into the computer, that many people will be raised above the poverty line (at no cost to Achilles) and the door will open. There is no other way of opening the door. Because of the mechanics of the machine, any door-opening number takes as much time and effort to enter (negligible) as any other.

Achilles enters a number (p) which, although fairly large, is significantly less than n . We ask him why he opted not to raise a further $n-p$ people above the poverty line. He replies that he is a Satisficing Consequentialist who thinks that saving p people from poverty in one day is "good enough". He thus sees no reason to save more people, and doesn't think he's done anything wrong.

Achilles' choice is morally unacceptable. If he really could have saved more people from poverty at absolutely no cost to himself, then he should have done so. His justification of his choice is, on the face of it, absurd.

It will be objected that I'm being unfair to Slote. My story is, of course, implausible, incomplete and, quite possibly, impossible. However, the Magic Game is effectively identical to one of Slote's own fairy stories:⁹⁴

The Dead Warrior : Eric, a great warrior, dies and arrives in Valhalla. The gods reward him by offering to grant any wish he makes. Eric asks that his family and their descendants be made "fairly well-off" for the rest of their lives. The gods ask him if he means as well-off as possible. "No", Eric replies, "I think fairly well-off would be good enough".

⁹⁴ibid., p. 26.

Slote endorses Eric's choice. He argues that Eric has done nothing wrong by not requesting that his family be made as well-off as possible. He does enough by ensuring that they will be fairly well-off.

However, there is another, more serious, dimension in which Eric satisfices, one which Slote ignores. Eric requests benefits *only* for his family. This is sub-optimal, as he could have considered everyone else. In particular, he could have asked that the many people who are (or will be) living in poverty or misery also be made fairly well-off. Once this dimension is noted, the similarity to Achilles' story is striking.

Slote's defence of Eric's choice is based on the following principle:⁹⁵

ordinary morality permits one sometimes to do less than the most or best one can do for others as long as what one does produces a large or sufficient amount of good for others.

Our discussion suggests that (in the context of Slote's own example) his notion of "large or sufficient amount of good for others" is ambiguous between the following two interpretations:

The Individual Interpretation : one does a large or sufficient amount of good for others if and only if one produces a large or sufficient amount of good *for each individual* who is in a position to benefit from one's actions.

⁹⁵ibid. p. 27.

The Total Interpretation : one does a large or sufficient amount of good for others if and only if the total amount of good one produces *for other people in general* is above a certain threshold.

In his discussion of Eric, Slote seems to adopt the Individual Interpretation. We don't blame Eric for not requesting additional benefits *for his family*, as he's already provided enough benefit *for them*. However, Eric's family are not the only people affected by his satisficing. Everyone else loses out too. Eric fails to do the best he can for many people for whom he hasn't produced any benefits at all.

Under the Individual Interpretation, for any given person, Eric is permitted to do less than the best he can for that person if and only if he's already produced a large or sufficient amount of good for that particular person. There are many people for whom this is not the case. Consider a person living in poverty and misery (call her Pauper). In making his choice, Eric could request a greater benefit for Pauper than he does. He's only allowed to do this if he's already providing a "large or sufficient amount of good" for Pauper. But Eric hasn't done any good for Pauper at all. So Eric's action will not be permitted under the Individual Interpretation.

In order to satisfy the Individual Interpretation, Eric would have to produce a large or sufficient amount of good for *each* individual. The following request would suffice: please make everyone (who will ever live) satisfactorily well-off. Unfortunately, this request would undermine the Undemanding Thesis. To establish that thesis, Slote must argue as follows: Eric is in a situation where he can (costlessly) do as much good as possible. In that situation, Eric acts on a satisficing maxim. That maxim is morally acceptable. Acting on that maxim would not be too demanding in our everyday lives. So morality need not be too demanding.

If Eric is to satisfy the Individual Interpretation, then he must be using the following maxim: You don't have to make everyone maximally well-off. It's enough to make everyone who will ever live fairly well-off. Once you've done that, you can get on with your own personal projects. Given the state of our world, following this maxim would, in practice, be every bit as demanding as Maximising Consequentialism. It would leave us no time at all for our personal projects, as there would always be someone who wasn't fairly well-off and whom we were in a position to help.

If we want to use Eric's story to establish the Undemanding Thesis, we must turn to the Total Interpretation. Under that interpretation, I'm allowed to refrain from (costlessly) providing the greatest possible benefit to a particular person even if I haven't provided that particular person with any benefit at all, so long as the total amount of good I've provided to other individuals is sufficiently great.

Unfortunately, proponents of the Total Interpretation must approve of Achilles' behaviour. To see this, let G be the amount of good beyond which I'm allowed to refrain from assisting people, let g be the amount of good which is produced by raising one person from poverty⁹⁶ and let m be the smallest whole number such that $mg > G$ (i.e., raising m people from poverty would produce at least G amount of good). So long as he makes p (the number of people he saves from poverty) at least equal to m , Achilles can thus ensure that he does enough good to satisfy the Total Interpretation.⁹⁷ He will then be absolved from providing any benefits to anyone else.⁹⁸

⁹⁶I've made the simplifying assumption that each person's rise from poverty has the same value, but nothing of substance turns on this.

⁹⁷ m will be significantly less than n . Otherwise, it would be impossible to do enough good without saving nearly everyone from poverty, which would undermine the Undemanding Thesis.

⁹⁸One disanalogy between the two stories as they stand is that Eric is able to choose which people will be saved, whereas Achilles is not. However, nothing turns on this, as we can easily add an additional feature to

Slote must endorse Achilles' choice. For Slote cannot establish the Undemanding Thesis without adopting the Total Interpretation, under which that choice is acceptable. Yet it seems absurd to sanction Achilles' behaviour. Achilles refuses to (costlessly) provide a vital good to a vast number of people. Those people's needs clearly provide Achilles with a reason for providing those goods. Achilles must respond either by producing a positive argument in favour of the action he performs, or by producing a negative argument demonstrating that he is permitted to ignore such a strong reason to act otherwise. Yet Achilles is unable to produce such reasons.

4.4: Slote's Moral Reasons

However, we should not be so hasty. After all, in analogous cases, Slote claimed to find a reason for blatant economic satisficing. Perhaps he will be able to find similar reasons in the moral case as well. To establish the Undemanding Thesis, Slote must demonstrate that there is a certain number (q) which satisfies the following two conditions:

- (i) saving q people from poverty is "morally good enough", even if one could have saved more than q people at no cost to oneself;
- (ii) saving q people from poverty would not place an unreasonable demand on my resources.

If we can find some q which meets both these requirements, then (by setting p equal to q) we can provide Achilles with a morally acceptable reason for his action.

Achilles' machine, allowing him to make that choice at no additional cost. This alteration would not render Achilles' refusal to save everyone from poverty morally acceptable!

In the economic case, Slote offered the following two reasons for blatant satisficing:⁹⁹

(1) *No Need* : the agent doesn't need more; or

(2) *Satisfaction* : the agent is perfectly satisfied with her present state.

We must now ask if analogous reasons can be found in the case of blatant moral satisficing. We begin with the argument from No Need. This seems to have (perhaps *inter alia*) the following moral analogues:

(1a) *No Agent Need* : the agent doesn't need more;

(1b) *No Dearest Need* : the agent's nearest and dearest don't need more;

(1c) *No Human Need* : no one at all needs more.

Agent Need and Dearest Need are both absent in Achilles' case. Neither he nor his family need anyone else to be saved. Unfortunately, such absence of need is not, in itself, a morally acceptable reason. Morality can't be just a matter of considering your own needs and those of your nearest and dearest. It also requires you to take account of the needs of others. You can't decline to (costlessly) provide a benefit for someone else simply because you don't need them to have that benefit: you must also ask if *they* need it.¹⁰⁰

⁹⁹Recall that these can be treated either as reasons to satisfice, or as reasons why satisficing is permitted (see the discussion in footnote 91 above)

¹⁰⁰I do not wish to deny that people are often permitted to refuse to provide certain benefits. I deny only that such permissions can arise simply because the agent doesn't need the benefit to be conferred. Of course, there may be moral systems which are constructed solely out of special, personal obligations. However, I don't think any plausible morality can give no weight at all to the needs of persons beyond the agent's circle of acquaintances.

The argument from No Human Need looks more promising. It is at least plausible to say that, once you've taken care of all the needs of everyone affected by your actions, your moral obligations are fully discharged. However, Human Need is not absent in Achilles' case. He leaves many people in poverty. People need to be saved from poverty. This argument thus cannot give Achilles a reason for not saving everyone from poverty.

Moreover, the argument from No Human Need could not provide the basis for any undemanding morality. It tells us that we can only pursue our own personal projects once every need of every person is met. In practice, this would render the Total Interpretation as demanding as the Individual Interpretation, as there will always be unmet needs which we ourselves could meet. No version of the argument from No Need can ground any morality which is both acceptable and not unduly demanding.

The argument from No Need cannot provide Achilles with a credible reason for his blatant moral satisficing. All the possible interpretations are either morally unacceptable or are not applicable to Achilles' situation. We turn now to Slote's second reason, the argument from Satisfaction. Once again, there are several moral analogues:

(2a) *Agent Satisfaction* : the agent is perfectly satisfied with the present state of the world;

(2b) *Dearest Satisfaction* : the agent's nearest and dearest are perfectly satisfied with the present state of the world;

(2c) *Universal Satisfaction* : everyone is perfectly satisfied with the present state of the world;

(2d) *Ideal Satisfaction* : an ideal moral observer would be perfectly satisfied with the present state of the world.

The first three arguments will fare no better than the analogous arguments concerning need. Agent Satisfaction and Dearest Satisfaction are not, in themselves, morally acceptable reasons. If Achilles and his family are satisfied with the world as he leaves it, then this merely reflects rather badly on them. On the other hand, Universal Satisfaction will clearly not be present in Achilles' case, as those who remain in poverty will be dissatisfied with the state in which he leaves the world. None of these arguments could be used to ground an acceptable morality which was not unduly demanding.

The argument from Ideal Satisfaction looks more promising. Perhaps I am only required to do enough to satisfy an ideal moral observer. However, an ideal moral observer would presumably not be satisfied with a situation in which many people were living in poverty. Such a state might be satisfactory if it were preferable to every available alternative. But Achilles' alternatives include situations in which no one is left in poverty. Compared with such alternatives, a situation in which some people remain in poverty would not satisfy an ideal moral observer.

Defenders of Achilles might reply that, while an ideal moral observer would not be satisfied with the situation in which Achilles leaves the world, such an observer might nonetheless be satisfied with what Achilles has done. Whatever the plausibility of this line of argument, it is incompatible with Slote's approach, as he seeks justifications for satisficing which are grounded in features of outcomes, not in features of agents. It must be the outcome, and not the action, which is satisfactory.¹⁰¹

¹⁰¹As the Ideal Satisfaction Argument is unacceptable on other grounds, I propose to ignore the standard problems regarding ideal observer theories in ethics.

Neither of Slote's reasons for economic satisficing has an acceptable moral analogue. In practice, the Total Interpretation will be no less exacting than the Individual Interpretation. Slote is thus unable to produce a morality which is significantly less demanding than Maximising Consequentialism.

4.5: Slote's Mistake

Slote may be able to find another justification for Achilles' action. However, it's hard to see what this might be. This is because Slote's position rests on a mistaken analogy between rationality and morals. In general, there are three separate domains to which the notion of "good enough" may be applied:

- (1) *outcome* : the situation which resulted once the action had been performed was "good enough";
- (2) *contribution* : the agent's contribution (the difference she made to the outcome) was "good enough";
- (3) *cost or effort* : the agent expended enough energy, paid a high enough price, ran enough of a risk, etc.

In discussions of economic rationality we usually speak of "good enough" results: "I was satisfied with that outcome, it was good enough". In his ethical discussion Slote appears to be focusing on the agent's contribution, but his reasons are more relevant to the outcome itself. If the fact that certain needs have been met is sufficient to absolve me from further contributions to the good, then, in a situation where those needs are already met before I

act, I don't have to make any contribution. If the situation is satisfactory before I start I don't have to do any good at all.

We could rewrite Slote's reasons in terms of contributions rather than outcomes.

Borrowing a distinction from Thomas Hurka,¹⁰² we can distinguish two ways in which an agent's contribution might be said to be "good enough". On the first ("absolute") interpretation, a contribution is good enough if and only if it is above a certain absolute threshold. This threshold would be the same for all agents and in all situations. On the second ("comparative") interpretation, an agent contributes enough if and only if she produces at least a certain proportion of the maximum contribution available to her in her situation.

Unfortunately, for Slote's purposes, neither notion of "good enough" contribution is any more promising than the notion of "good enough" outcome. To see this, we need only return to Achilles' story. There is no threshold beyond which Achilles has a legitimate reason not to (costlessly) rescue more people from poverty. Similarly, there is no percentage of his maximum contribution beyond which Achilles has a legitimate reason not to (costlessly) go.

Slote's reasons fail because the notion of "good enough" which gives his position its intuitive appeal is not one related to either outcomes or contributions. Rather, it is the notion encapsulated by such turns of phrase as: "she should have a rest, she's done enough good for one day", "we couldn't expect any greater sacrifice - it was a good enough effort". Yet, here, the "good enough" refers, not to the agent's contribution, nor to the outcome resulting from that contribution, but to the effort required to produce that outcome.

¹⁰²See Hurka 1990.

In other words, "good enough" is used in sense (3), not in either of senses (1) and (2). To see this, consider the following story:

Two Doctors : Speedy and Steady are two doctors working in an over-stretched emergency ward. They both start work at time t when many patients require urgent treatment. Steady tires easily and is not a fast worker. At time $t+1$, Steady is exhausted and takes a rest having treated m patients, leaving many patients requiring treatment. Speedy is extremely fit and works fast. At time $t+2$, Speedy has treated $2m$ patients and is still feeling refreshed. However, Speedy takes a rest anyway, leaving behind a few patients who require treatment.

I suggest that, in this situation, Steady's decision is justified whereas Speedy's is not. Steady stops because she is exhausted and needs a rest. Speedy takes a rest without needing one, leaving vital work undone. However, if we focus on either outcomes or contributions, then we cannot reach this conclusion. The overall situation is better at $t+2$ than at $t+1$, while in absolute terms Speedy's contribution is greater than Steady's. We may also assume that, because she stops when only a few patients require treatment, Speedy's contribution is higher as a percentage of the maximum contribution available to her than is Steady's. We thus cannot approve of Steady's rest without also approving of Speedy's rest. This is unacceptable.

When assessing Steady and Speedy, we don't think that a doctor is entitled to a rest simply because she's cured n people today or because n people have been cured today, irrespective of whether or not she needs a rest. Rather, we think that she's entitled to a rest because she's expended y amount of energy in healing people (irrespective of how many people she's healed or of how many people have been healed overall). It is the effort, not the result, which is good enough. Similarly, when someone pays a "high enough" price, we

excuse them from further danger because it would be unreasonable to expect a greater sacrifice. The question of how much good that sacrifice achieved is of secondary importance.¹⁰³

It is tempting to think that Slote can construct a Satisficing Consequentialism based on the notion of "good enough effort". However, this move would undermine Slote's reply to the Demandingness Objection. Slote seeks to show that morality doesn't place unreasonable demands upon moral agents. His original theory provides an explanation: morality doesn't place unreasonable demands upon agents because it never requires them to produce more than a certain amount of good. Here, the limit on the costs an agent can be asked to bear is explained by appealing to a feature of the outcome of her actions. This is both an *explanation* and a *Consequentialist* explanation. If we turn instead to the notion of "good enough" effort (or "high enough" cost), our reply to the Demandingness Objection will be as follows: morality doesn't place unreasonable demands upon agents because it never requires them to make too great a sacrifice. We might as well say that morality doesn't place unreasonable demands upon agents because morality never makes unreasonable demands. This is not an explanation. Nor is the resulting theory genuinely Consequentialist, as it defines moral obligations in terms of costs to the agent, not just in terms of features of outcomes. Replacing "good enough outcome" with "good enough effort" is not a plausible move for the Satisficing Consequentialist.

¹⁰³Of course, the degree of effort involved cannot be the only relevant factor, as sacrifices which are not normally required may become necessary when particularly urgent consequences are at stake. For instance, a given level of sacrifice might be required to save another person's life, but not to save them from inconvenience.

4.6: Conclusions

Even if blatant economic satisficing is occasionally rational, an act is never acceptable simply because it is an instance of blatant moral satisficing. Consequentialists seeking to rebut the Demandingness Objection must look elsewhere. Our discussion suggests that they should concentrate on features of the agent (her efforts or the costs of her actions) and not just on features of outcomes (either the outcome itself or the agent's contribution to that outcome). However, Consequentialists must also be wary of straying too far from their roots. Their real challenge will be to do full justice to agent-centred features without effectively abandoning Consequentialism altogether. Our discussion suggests that this will not be easy. For instance, Collective Consequentialists (and others who seek to rebut the Demandingness Objection by departing from traditional Act Consequentialism) may face the same dilemma as Satisficing Consequentialists. Appeals to features of outcomes cannot provide an acceptable response to the objection, whereas appeals to features of the agent fit uneasily within a broadly Consequentialist theory.

The notion of cost to the agent recurs throughout the remainder of the thesis, as both Scheffler's Hybrid View (discussed in Chapter FIVE) and my own tentative theory (see especially Chapter SEVEN) attempt to make more room for the moral significance of that notion than it is granted by traditional Consequentialism.

Chapter FIVE: Scheffler's Agent-Centred Prerogative

Following our rejection of Slote's Satisficing Consequentialism, we now consider a form of Consequentialism which explicitly recognises the moral significance of the cost to the agent.

Samuel Scheffler has outlined a "hybrid view" of ethics.¹⁰⁴ Someone might propose this Hybrid View as a solution to the Demandingness Objection. Such a person would have three principal aims:¹⁰⁵

- (1) To show that agents are sometimes *allowed* to not perform the act with the best outcome;
- (2) To show that (1) does not imply that agents are ever *required* not to perform the act with the best outcome;
- (3) To achieve (1) and (2) within an essentially Consequentialist theory.

In particular, the Hybrid Theorist wishes to accomplish both (1) and (2) without appealing to two common non-Consequentialist distinctions: the distinction between "doing x" and "allowing x to happen", and the distinction between "intending x (either as an end or as a means)" and merely "foreseeing x".

¹⁰⁴Scheffler 1982.

¹⁰⁵It is not clear that Scheffler himself would make quite this response to the Demandingness Objection. In particular, in his more recent book (Scheffler 1992c), Scheffler's defence of a "moderate" moral theory operates on a more abstract level.

5.1: Outlining the Hybrid View

We begin by outlining the Hybrid View. Under Consequentialism, the weight an agent is allowed to give to her own personal projects must be in strict proportion to their impersonal value. For instance, I should only pursue a particular hobby or interest of my own to the extent that the well-being I receive from so doing is at least as great as the total well-being I could generate for others by acting differently. The basis of the Hybrid View is the agent-centred prerogative, which "has the effect of denying that one is always required to produce the best overall states of affairs".¹⁰⁶ Scheffler goes on to say that "a plausible agent-centred prerogative would allow each agent to assign certain proportionately greater weight to his own interests than to the interests of other people".¹⁰⁷ The justification for such a prerogative is that it "constitutes a structural feature whose incorporation into a moral conception embodies a rational strategy for taking account of personal independence, given one construal of the importance of that aspect of persons".¹⁰⁸

We must now ask how the Hybrid View is related to the other forms of Consequentialism we have discussed. The Hybrid View is closer to Individual than to Collective Consequentialism, as it assesses acts on the basis of their individual consequences. The Hybrid View is act-focused, rather than desire or motive or rule focused, as it assesses acts directly. The Hybrid View has some affinities with Slote's Satisficing Consequentialism, as it attempts to provide the agent with some protected space within which to pursue her own projects at the expense of the general good. Unlike Slote, however, the Hybrid View makes room for the agent by explicitly appealing to the value her projects have for her. It therefore looks as if the Hybrid View will fare better against the Demandingness Objection than Satisficing Consequentialism did. In particular, stories such as Achilles Tale pose no

¹⁰⁶Scheffler 1982, p. 5.

¹⁰⁷ibid., p. 20.

¹⁰⁸ibid., p. 67.

threat to the Hybrid View. Recall that Achilles would have borne no cost at all if he had produced the greatest possible good. However much weighting he is allowed to give to his own interests, a zero cost can never outweigh the forsaken goods. So Achilles cannot claim that he is exercising an agent-centred prerogative.

5.1.1: Three Interpretations

As Scheffler initially presents it, the notion of the agent-centred prerogative is somewhat vague. Three obvious interpretations suggest themselves.¹⁰⁹ We begin by contrasting two of these. In the first interpretation, we contrast two assessments of any outcome. Under the Impersonal Assessment, the value of an outcome is the sum of the utilities of all the agents in that outcome.¹¹⁰ We then assess outcomes from the personal viewpoint of the individual agent (call her A). On A's Personal Assessment, the "weighted value" of an outcome will be equal to the sum of the utilities of all the agents in that outcome apart from A, plus A's utility in that outcome multiplied by a particular factor (call this factor the disproportionate weighting which A is allowed to give to her own utility).

The agent's own utility is thus given greater weight in the personal assessment than in the impersonal one.¹¹¹ The agent-centred prerogative will then allow an agent to produce a sub-optimal outcome so long as the weighted value of that outcome is greater than the impersonal value of the optimal outcome. If y is the act with the best available outcome, then we can say that A is permitted to perform another act (x) if and only if A's utility if she does x , when multiplied by the disproportionate weighting of A's utility, is greater than the

¹⁰⁹Scheffler himself does not discuss the precise interpretation of his agent-centred prerogative. Perhaps he would reject any "mathematical" interpretation of the notion of "disproportionate weight". However, his use of that notion seems to suggest that some such interpretation is appropriate.

¹¹⁰For simplicity's sake, I shall assume a straightforward Utilitarian account of the value of outcomes. This assumption is not essential either to Scheffler's theory or to my criticisms of it.

¹¹¹Individual Consequentialism is the limiting case in which the factor by which the agent's own utility is multiplied is 1, so that the personal and impersonal assessments will be identical.

difference between the total (impersonal) value of the outcome which would result from A's doing y and the total (impersonal) value of the outcome which would result from A's doing x.

On a second interpretation, an agent (A) is permitted to produce a sub-optimal outcome if and only if the cost she would have incurred in producing instead the optimal outcome (that is, the difference between A's utility if she does x and her utility if she does y), when multiplied by a certain set factor (call this factor the disproportionate weighting of the cost to A), is greater than the difference between the total (impersonal) value of the outcome which would result from A's doing y and the total (impersonal) value of the outcome which would result from A's doing x.

These two interpretations are not equivalent. In particular, the second interpretation allows the agent to give disproportionate weight to the *cost* she would have borne in producing the optimal outcome, whereas the first interpretation allows the agent to give disproportionate weight to her *utility*. The difference is between an emphasis on comparative levels of utility and an emphasis on absolute levels of utility. As a result, our first interpretation has the unfortunate consequence that those with higher utility will have wider prerogatives than those with lower utility. Whether or not a given sacrifice is required will depend on how well-off the agent is initially, not just on the level of sacrifice involved. To see this, consider the following situation, in which two agents each face a choice between two options (where option y will produce more value overall than option x):

Agent	Utility if does x	Utility if does y
A	100	50
B	200	190

Now, under our first interpretation of the agent-centred prerogative, B is more likely to be allowed to do x rather than y than A is. This is because B's level of utility if she does x will be higher than A's level of utility if she does x. This is despite the fact that the cost to A of doing y rather than x is much greater than the cost to B of doing y rather than x. That is, A would be required to sacrifice more utility to bring about the optimal outcome than B would. On our second interpretation, by contrast, A is more likely to be allowed to do x than B is, precisely because the sacrifice which would be demanded by forbidding her from doing x would be far greater than the corresponding sacrifice imposed on B.

This consequence of our first interpretations is clearly objectionable. Therefore, we should prefer a version of the agent-centred prerogative based on the disproportionate weighting of comparative, not of absolute, utilities. Our second interpretation is preferable to the first.¹¹²

This second interpretation must be contrasted with one suggested by Scheffler himself in a latter article, where he says:¹¹³

Suppose, in other words, that each agent were allowed to give *M* times more weight to his own interests than to the interests of anyone else. This would mean that an agent was permitted to perform his preferred act (call it *P*), provided that there was no alternative *A* open to him, such that (1) *A* would produce a better

¹¹²In a footnote in his book, Scheffler seems to suggest that the first of these two interpretations is more demanding than the second (Scheffler 1982, p. 20). This is not strictly correct, as the demandingness of the Hybrid View on either interpretation depends entirely on the constant factor by which the agent's cost or utility is multiplied. Nothing in the basic structure of either interpretation tells us how demanding the agent-centred prerogative will be. What is true is that the two interpretations give different accounts of the comparative demands on different agents. In particular, on the first interpretation, agents with different absolute levels of utility will face different demands, while this is not necessarily the case under the second interpretation.

¹¹³Scheffler 1992a, p. 378.

overall outcome than P, as judged from an impersonal standpoint which gives equal weight to everyone's interests, and (2) the total net loss to others of his doing P rather than A was more than M times as great as the net loss to him of doing A rather than P.

The difference between our second interpretation and Scheffler's is simple enough. Whereas we have been comparing the agent's preferred act only with the act which would produce the best consequences overall (as judged from an impersonal standpoint), Scheffler compares the preferred act with every alternative act available to the agent at that time.

To see how these two interpretations can come apart, consider the following story. An agent (A) has three options: donate all her money to charity, donate most of her money to charity, or donate nothing. The more A gives away, the greater will be the cost to her of making a further donation, as she will be required to give up progressively more central elements of her lifestyle. The values and costs of her options are represented in the table below:

	Impersonal Value of Outcome	Cost A would face
Donate All	100	10
Donate Most	99	1
Donate Nothing	50	0

Let us assume that, on either interpretation of the agent-centred prerogative, each agent is allowed to give six times more weight to his own interests than to the interests of anyone else. On our tentative interpretation, A is permitted to donate nothing. This is because the difference between the amount of value which would result from her donating all and the amount of value which would result from her donating nothing is only five times as great as the difference between the cost to A of donating all and the cost to A of donating nothing. As five is less than six, A is allowed to donate nothing rather than donating all. As donating all is the act with the best consequences, my interpretation says that A has an agent-centred prerogative to donate nothing.

By contrast, on Scheffler's interpretation, A is not permitted to donate nothing. This is because the difference between the amount of value which would result from her donating most and the amount of value which would result from her donating nothing is forty-nine times as great as the difference between the cost to A of donating most and the cost to A of donating nothing. As forty-nine is greater than six, this means that A is not permitted to donate nothing rather than donating most. On Scheffler's interpretation, this is enough to demonstrate that A does not have an agent-centred prerogative to donate nothing.

However, this does not mean that A is required to donate all. The difference between the amount of value which would result from her donating all and the amount of value which would result from her donating most is less than the difference between the cost to A of donating all and the cost to A of donating most. This means that A is permitted to donate most rather than donating all. Yet donating all is the only option which would produce better consequences than donating most. So A has an agent-centred prerogative to donate most.

In general, the two interpretations will diverge in cases where the agent faces increasing marginal costs; that is, where the cost to the agent of producing each subsequent increment of goodness increases.

We must now choose between these two interpretations. We should note that both are preferable to our first interpretation, as both deal with comparative rather than absolute levels of utility. We should also note that we are interested in the Hybrid View as a general approach, not in Scheffler's personal views *per se*. Scheffler's own interpretation should not be regarded as authoritative.

However, it still seems to me that Scheffler's interpretation is preferable. Both interpretations agree that A is not required to donate all her money to charity. They disagree over the question of whether or not she is allowed to donate nothing rather than donating most. On this question, our tentative interpretation seems implausible, as it permits A to fail to bring about a very great amount of good, even though she could do so at almost no cost to herself. Indeed, as this interpretation takes no account whatever of the cost to A of donating most rather than nothing, we can make that cost vanishingly small, and still conclude that A is not required to produce the additional good. This seems to contradict the conclusion we reached in our discussion of Slote. If an agent can produce a significant amount of good at negligible cost to herself, then she should be required to do so.

Scheffler's option-by-option comparison seems to yield the right result in those situations where it differs from our tentative interpretation. Accordingly, the former is to be preferred. All future references to the agent-centred prerogative or the Hybrid View are to Scheffler's interpretation, unless otherwise stated.

5.1.2: Applying the Hybrid View to Affluent's Tale

It may seem odd that I have spent so much time concluding that Scheffler's own interpretation of the agent-centred prerogative is the best. However, it is important to be very clear about the precise structure of the Hybrid View. This is partly because the different interpretations we have discussed are often not adequately distinguished. More importantly, the differences between them are especially significant in the very type of situation around which our discussion revolves, as the story we used to distinguish them is merely a simplified version of Affluent's Tale.

Let us now apply the Hybrid View to a more realistic retelling of Affluent's Tale. We begin with a number of simplifying assumptions. Assume that Affluent has available to her a set number of dollars (call it n), and that she can donate to Oxfam any number of dollars from zero to n . Given the size of Oxfam's operations as compared to Affluent's resources, we can assume that the marginal good produced by each additional dollar donated is constant. In other words, the difference between the amount of good produced by a donation of x dollars and that produced by a donation of $x+1$ dollars will be the same for any value of x between zero and n . By contrast, we can assume that the marginal cost to Affluent of each additional donation of one dollar is increasing. That is, the cost to Affluent of donating an additional dollar once she's already given most of her income will be much greater than the cost to her of donating an additional dollar when she hasn't given anything as yet.

Under Scheffler's interpretation of the agent-centred prerogative, we can prove that, unless Affluent is required to donate all her money to charity, there is some amount of dollars between zero and n (call it x), such that (i) Affluent is permitted to donate x dollars to charity rather than donating more, (ii) Affluent is required to donate at least x dollars to charity, and (iii) the additional amount of good which would be produced by a donation of

$x+1$ dollars rather than one of x dollars is equal to the cost to Affluent of donating her $(x+1)$ th dollar multiplied by the extra weight Affluent is permitted to give to her own interests over the interests of anyone else (Scheffler's M).¹¹⁴

These results depend upon the assumption of increasing marginal cost. However, if we reject increasing marginal cost, then it must be the case either that Affluent is permitted to donate nothing at all or that she is required to give all of her money to Oxfam. The possibility that Affluent is required to make some donation, but not required to bring about the best possible consequences, will be ruled out.¹¹⁵ Yet it is precisely this possibility that proponents of the Hybrid View seek to defend. Accordingly, the assumption of increasing marginal cost must be retained.

¹¹⁴The proof of this is as follows: If Affluent is not required to donate all her money to Oxfam, then there must be at least one amount of money (y) such that Affluent is permitted to donate y dollars to charity rather than donating any more. Let x be the smallest amount of money such that x meets condition (i). It follows that, for any amount of money less than x , Affluent is not allowed to donate that amount of money rather than donating more. If Affluent is not allowed to keep all her money for herself, then x must be greater than zero. So x meets condition (ii) as well as condition (i). Now, if Affluent is permitted to donate x dollars, then she's permitted to donate x dollars rather than donating $x+1$ dollars. The marginal utility of that next dollar to Affluent, when multiplied by the extra weight Affluent is permitted to give to her own interests (call the result the weighted cost), must be at least as great as the marginal value of a dollar to Oxfam. Assume for the moment that, at x , the weighted cost is greater than the marginal value. Recall that both the weighting and the marginal value are constant, whereas the marginal cost is increasing. Therefore, there must be some amount of money (z) less than x , such that the weighted cost to Affluent of donating z dollars is equal to the marginal value of an extra dollar. Affluent would then be permitted to donate z dollars rather than donating $z+1$. However, for each dollar we add beyond $z+1$, the marginal value remains constant, whereas the weighted cost increases. Therefore, if Affluent is permitted to donate z rather than $z+1$, she will be permitted to donate z rather than any amount greater than z . But this would contradict our assumption that x is the least amount of money such that Affluent is permitted to donate no more than that amount. So, by reductio, we can conclude that, at x , the weighted cost is not greater than the marginal value. Therefore, the weighted cost must be equal to the marginal value. So x meets condition (iii). QED. (For this proof, I made the simplifying assumption that money comes in discreet amounts, which I identify with dollars.)

¹¹⁵The proof is straightforward. Both marginal value and the weighting the agent is allowed to give to her own interests are constant. If the marginal cost to the agent of donating an extra dollar is also constant, then the relationship between weighted cost and marginal value will be constant. Either weighted cost is always at least as great as marginal value, in which case Affluent is permitted to donate any amount she chooses rather than donating more, or weighted cost is always less than marginal value, in which case Affluent is always required to donate more if she is able to. Either a zero donation is permitted or a total donation is required.

The Hybrid Theorist's reply to the Demandingness Objection should now be obvious. By altering the value of M (the weighting the agent is allowed to give to her own interests), we can change the value of x (the amount of money she is required to donate). By choosing an appropriate weighting we can thus ensure that the maximum donation Affluent is required to make is not unduly demanding.

5.2: Objections to the Hybrid View

I propose to discuss two objections to Scheffler's Hybrid View. The first (the Wrong Facts Objection) seeks to show directly that the Hybrid View is not an acceptable response to the Demandingness Objection, while the second (Kagan's objection that Scheffler cannot account for the distinction between doing and allowing) is a more general complaint against the Hybrid View, which does not bear directly on the question of demandingness itself.

5.2.1: The Wrong Facts Objection

Scheffler's theory is open to a version of the Wrong Facts Objection, which arises as follows:¹¹⁶

Affluent's Fourth Ignorance : Affluent is sitting at her desk with her cheque book. She has heard a rumour that the effectiveness of Oxfam's operations (that is, the amount of good they can do for each dollar they receive) has been drastically altered. However, she does not know whether or not this rumour is true; or, even if it is true, what the nature of the alteration has been. To simplify, let us say that there are three possibilities

¹¹⁶Other versions of this objection have been discussed in detail in chapters TWO (section 2.2.1) and THREE (section 3.1) above.

Inefficient Oxfam : Oxfam's efficiency has declined to 10% of its previous level. That is, each dollar spent now produces only one tenth as much good as it would have previously.

Normal Oxfam : Oxfam's efficiency is unchanged.

Superefficient Oxfam : the efficiency of famine relief operations has increased tenfold. Each dollar spent now produces ten times as much good as it would have previously.

We now apply our interpretation of Scheffler's agent-centred prerogative to this story. Let us assume, as we did in Chapters TWO and THREE, that in Normal Oxfam the maximum required sacrifice is 10% of Affluent's income. That is, Affluent has given 10% of her income to Oxfam when she reaches the point at which the amount of additional utility she could gain by spending her next dollar on herself, when multiplied by the disproportionate weighting she's allowed to give to her own costs, is equal to the amount of additional value Oxfam could produce (in Normal Oxfam) by spending that dollar.

Now, we know that, in Superefficient Oxfam, the amount of additional value Oxfam could produce by spending any given dollar is ten times as great as the amount of additional value they could have produced with that same dollar in Normal Oxfam. Therefore, in Superefficient Oxfam, Affluent must continue making donations until the amount of additional utility she would gain by spending her next dollar on herself is ten times as great as it would have been at the point at which she stopped making donations in Normal Oxfam.¹¹⁷ In other words, Affluent must keep donating money until the sacrifice from giving an extra dollar is ten times as great as the sacrifice from giving an extra dollar when her current level of donation was 10%. To put it crudely, in Superefficient Oxfam Affluent

¹¹⁷This follows trivially from our definition of the agent-centred prerogative.

is required to give up things which are ten times as valuable to her as the things she is required to give up in Normal Oxfam.

Similarly, we know that, in Inefficient Oxfam, the amount of additional value Oxfam could produce by spending any given dollar is only one tenth as great as the amount of additional value they could have produced with that same dollar in Normal Oxfam. Therefore, in Inefficient Oxfam, Affluent is only required to continue making donations until the amount of additional utility she would gain by spending her next dollar on herself is one tenth as great as it would have been at the point at which she stopped making donations in Normal Oxfam.¹¹⁸ To put it crudely, in Inefficient Oxfam the agent-centred prerogative would allow Affluent to indulge in things which are only one tenth as valuable to her as the things she would be allowed to indulge in in Normal Oxfam.¹¹⁹

In Superefficient Oxfam, many actions will be forbidden which would have been permitted in Normal Oxfam. In Inefficient Oxfam, many actions will be permitted which would have been forbidden in Normal Oxfam. For instance, consider an activity which provides Affluent with nine times as much utility, for each dollar she spends, as her children's music lessons. In Normal Oxfam this activity would clearly have been permitted. In Superefficient Oxfam it will be forbidden. Similarly, consider another activity that provides Affluent with only one ninth of the utility per dollar that she derives from her children's music lessons. This activity will be permitted in Inefficient Oxfam, though it would clearly have been forbidden in Normal Oxfam.

¹¹⁸This follows trivially from our definition of the agent-centred prerogative.

¹¹⁹We should note that my claim here is not that Affluent will be required to give ten times as much money in Superefficient Oxfam as in Normal Oxfam, nor that she will be required in Inefficient Oxfam to give only one tenth as much money. Indeed, as a result of diminishing marginal utilities, the differences between the amounts of money she's required to sacrifice are likely to be less than factors of ten. However, the variations in the amounts of money Affluent is required to sacrifice will still be very great. Greater, I would suggest, than we can accept.

For any possible degree of sacrifice there will be a version of Superefficient Oxfam in which that degree of sacrifice is required of Affluent. For instance, we could retell the Superefficient Oxfam story so that the increase in Oxfam's efficiency was by a factor of one hundred rather than by a factor of ten. In that case, Affluent would be required to continue donating until the marginal utility of an extra dollar was one hundred times as great as the marginal utility per dollar spent on her children's music lessons. In a more extreme version of Affluent's Fourth Ignorance, she would literally have no idea at all how much (or how little) she ought to give to Oxfam. It seems unreasonable for Affluent to be made so totally a hostage to fortune.

Affluent's obligations and prerogatives depend crucially on which of the three situations she is in. Affluent must resolve her ignorance before she can have any idea at all what she is permitted to do. My objection is not that the difference between these stories should make no difference at all to the degree of Affluent's obligations. Rather, it is that those differences should not affect those obligations to the extreme degree which they do on Scheffler's account. In each of the three stories, the amount of good that Affluent foregoes by acting sub-optimally is much greater than the sacrifice she will be required to make by acting optimally. How much sacrifice she is required to make should not depend to so great an extent on precisely how much greater the loss in good is than the level of sacrifice. Like Hooker and Murphy, Scheffler seems to make the wrong empirical questions count for too much.

We now consider some replies that defenders of the Hybrid View might make to the Wrong Facts Objection.¹²⁰ The most obvious reply is to reject our attempts to formulate the agent-centred prerogative with any mathematical precision. Scheffler himself suggests that such

¹²⁰Some of these replies are similar to those available to Hooker and Murphy, which were discussed in Chapters TWO (section 2.2.1) and THREE (section 3.1).

formulations give the Hybrid View an implausible air of precision, and that no such formulation can be essential to the view itself.¹²¹ However, if we are given no formulation at all, then we have little idea what the Hybrid View really amounts to. The Hybrid View must be presented with some precision, even if it lacks fully mathematical precision. We should concentrate on alternative formulations of the Hybrid View, to see if any of these is able to avoid the Wrong Facts Objection.

The Wrong Facts Objection arises because M (the amount of extra weight an agent is allowed to give to her own interests when compared to the interests of others) is held constant. We could thus avoid the objection by allowing M to vary as the amount of value at stake varies. At the extreme, we could tie M to the difference between the value produced by making the best available donation, and the value produced by making a small donation. This revised Hybrid View would then give us exactly the same answers in each different story in the Wrong Facts Objection, thus removing the objection.

This extreme view is not acceptable. It amounts, in effect, to saying that there is a certain level of sacrifice which is such that Affluent is never required to make any sacrifice greater than that level, irrespective of how much or how little value is at stake. If a vast amount of value is at stake, then Affluent is still only required to make a moderate sacrifice. If virtually no value is at stake, Affluent will still be required to make the same moderate sacrifice. This reformulated Hybrid View will ask no more of Affluent to save the world than it would ask of her to save a single stranger from a mild headache. This position does not seem plausible.

We could avoid this consequence by adopting a less extreme version of this reformulation, one on which the amount of extra weight an agent is allowed to give to her own interests

¹²¹See Scheffler 1992a, p. 378.

when compared to the interests of others increases as the amount of value of stake increases, but where the former does not increase by as much as the latter. This middle-road would seem to avoid both the original Wrong Facts Objection and the problems raised in the preceding paragraph, as the level of sacrifice required of Affluent would increase as the amount of value at stake increased. Two forms of the Hybrid View which are similar in many ways to this last suggestion are discussed in detail in Chapter SEVEN, as they utilise the theoretical innovations developed in Chapter SIX.

A similar response to the Wrong Facts Objection would be to appeal to the idea that there are some sacrifices which can never be required. This would enable us to deny that Affluent is required to make much greater sacrifices in those cases where the amount of good she is able to produce is greatly increased. Briefly, I see three main problems with this response. The first is that it is not obvious that, in fact, there are any financial or material sacrifices so great that they are never required. Secondly, even if we admit that there are such sacrifices, we must admit that they are very demanding. For instance, it would surely be implausible to claim that there could be no situation in which any agent was required to give up most of her wealth for the sake of some greater good. Yet the principal claim of the Wrong Facts Objection was not that there was no limit to the sacrifice potentially required of Affluent under the Hybrid View, but rather that the Hybrid View would require a significant sacrifice of Affluent under certain circumstances. Unless we can rule out all significant sacrifices, we will be unable to respond effectively to this claim. A third problem is that setting an upper limit to the sacrifices which can be demanded of an agent can only possibly provide a solution to one half of the Wrong Facts Objection. The objection that, if the amount of value at stake is dramatically reduced, then too little will be required of Affluent, remains untouched.¹²²

¹²²We could, of course, provide a parallel response, to the effect that there is a certain minimum sacrifice which is always required of any agent in any situation. (For instance, perhaps we should always "give til it hurts".) However, such a view seems less plausible than Scheffler's original upper limit on sacrifices.

The problem faced by this response to the Wrong Facts Objection is that although the intuitions which underlie the rationale for the Hybrid View also provide support for the view that different levels of sacrifice are required in different circumstances, the appropriate differentiation of circumstances is very hard to capture within the Hybrid View. As we saw in Chapter FOUR, Slote's Satisficing Consequentialism fails because it requires the level of sacrifice required of an agent to be a function only of the amount of good the agent has done. Scheffler's Hybrid View runs into trouble because the level of sacrifice required of an agent is only allowed to be a function of the amount of good at stake and of the cost to the agent.

Another line of reply for proponents of the Hybrid View would be to argue that, so long as the Hybrid View is appropriately demanding in the actual world, it doesn't matter how demanding it would have been in other possible situations. The principal objection to this argument is that, when judging a moral theory, we are not only interested in what it prescribes in the actual world. It is not enough for a theory to happen to give the right results, it must give the right results for the right reasons. Part of what it means to give the right results for the right reasons is that the theory would have given the right results in certain possible but not actual situations. We should note also that the situations discussed in Affluent's Fourth Ignorance are by no means bizarre. They involve none of the alien invasions, miracle technologies or supernatural forces often found in the tales of philosophers. If a theory cannot give a right answer in a plausible counterfactual situation, then it must be rejected.¹²³

Another problem with this response is that not all situations in the actual world are like Affluent's situation. The Hybrid View may only be able to get the right result in Affluent's

¹²³A identical line of argument is discussed in section 2.2.1 in Chapter TWO.

case by giving the wrong answer in other cases in the actual world. This possibility is exploited by Kagan's objection, which is discussed in the next section.

5.2.2: Kagan's Objection

Kagan has argued that "an agent-centred prerogative not only permits agents to *allow* harm, it will also permit agents to *do* harm in the pursuit of their non-optimal projects".¹²⁴ Kagan cites two situations where the pursuit of my projects requires a large sum of money. In the first case, I don't have enough money myself, so I kill my uncle in order to inherit \$10,000. In the second case, I already have \$10,000, so I elect to spend it on my own projects rather than give it to a charity which would have used it to save a stranger's life. An agent-centred prerogative can permit acts of killing if and only if it also permits acts of letting die. If agent-centred prerogatives are to be any use at all, they must (at least sometimes) allow me to spend my money on myself rather than on saving the lives of others. Some acts of letting die must be allowed. So some acts of killing must also be accepted. Scheffler must permit agents to kill in order to advance their own personal projects at the expense of the general good.

Scheffler describes this alleged feature of the Hybrid View as the "symmetry condition", which arises because:¹²⁵

it seems to follow [from the Hybrid View] that if it is permissible for a person to allow an n-sized harm to befall someone else in order to avoid a q-sized cost to himself, then it must, everything else equal, also be permissible for the person to inflict an n-sized harm directly in order to avoid a q-sized cost to himself.

¹²⁴Kagan 1984, p. 251.

¹²⁵Scheffler 1992a, p. 380.

Act Consequentialism also allows some killing. However, Consequentialists have the defence that they only ever allow me to kill when this is necessary to bring about the best outcome. This response is unavailable to Scheffler, as he permits killing when it produces a worse outcome than not killing.

A response to Kagan's objection from within the Hybrid View will seek ways to differentiate acts of doing from acts of allowing by appealing either to the amount of value produced by different acts, or to the costs to the agent of performing them. Scheffler himself mounts just such a defence in two separate places: his 1982 book and his 1992 article.¹²⁶ We begin with the former discussion.

In his book, Scheffler admits that an agent-centred prerogative cannot, of itself, distinguish between various ways of pursuing personal projects. However, he argues that, in practice, there will be a difference. Doing harm is an act of commission rather than omission. It thus takes time and energy which cannot then be used in the pursuit of personal projects. By contrast, an act of allowing harm saves the agent time and energy, which can then be used in the pursuit of personal goals. Acts justified by agent-centred prerogatives will tend to be acts of allowing harm rather than doing harm.

Scheffler gives the example of a starving inner-city inhabitant. By allowing them to starve, I am able to stay at home to work on my own personal projects. So this act may be justified by an agent-centred prerogative. Yet taking active steps to kill them myself would leave me with even less, rather than more, time to devote to my projects. This act will thus not be justified by an agent-centred prerogative.

¹²⁶See Scheffler 1982, p. 23, and Scheffler 1992a, pp. 380-387.

Scheffler's argument will work in many cases. Perhaps most of those acts involving harm to others which are justified by agent-centred prerogatives will be acts of omission rather than commission. However, as Scheffler himself admits, it is implausible to claim that they all are. Situations are bound to arise in which pursuit of my personal projects is best achieved by some active doing of harm. Furthermore, it is not necessarily true that performing an act of commission limits the number of other options the agent is able to perform. For instance, we might imagine a case in which a particular act of harming another person was the only way for a given agent to open up for herself a whole range of otherwise unavailable options. Indeed, Kagan's example seems to be just such a case.

Scheffler's 1982 argument thus establishes only a very partial, contingent distinction between doing and allowing. The 1992 argument seeks to ground a much more substantial distinction. Focusing on Kagan's example of leaving a stranger to die versus killing my uncle, Scheffler appeals to four differences between the two acts.¹²⁷ These are:

- (a) the occurrence of a killing is not the same as the occurrence of a death of some unspecified kind;
- (b) the death of an uncle is not the same as the death of a complete stranger;
- (c) costs to the agent cannot be measured in purely financial terms. The non-financial cost to the agent of preventing a far-off death is much greater than the non-financial cost of not bringing about far-off deaths;
- (d) the costs to people other than the agent of preventing a far-off death are much greater than the costs to people other than the agent of not bringing about far-off deaths.

¹²⁷Scheffler 1992a, p. 380.

The last two differences form the crux of Scheffler's argument. Scheffler mentions the first two differences in order to suggest that Kagan's original example is not fair to the Hybrid View, as we are not really comparing two comparable choices. Scheffler suggests that a fairer comparison would be the following: "it will apparently be permissible to kill my rich uncle in order to inherit \$10,000 [as] most of us believe we would not be required to pay \$10,000 to prevent someone from murdering our uncle".

Scheffler notes that, in this more symmetrical comparison, Kagan's claim may have less force, as it is not clear that most of us believe that we would not be required to pay \$10,000 to prevent someone from murdering our uncle. This shows that at least some of the appeal of Kagan's original example derives from its illegitimate asymmetries. However, Scheffler does not press this point, as he admits that Kagan may well be able to produce a suitably symmetrical example in defence of his claim.

This brings us to the first of Scheffler's two main distinctions: the claim that the non-financial costs of killing are much greater than the non-financial costs of letting someone die. Scheffler lists two principal ways in which the psychological costs of killing differ from those of letting someone die.¹²⁸ The first is "the general empirical observation that the psychological costs of surrendering a benefit one already has tend to be greater than the psychological costs of doing without a comparable benefit that one has not yet acquired". The second is that if the money involved is "obtainable only by killing", then the agent will only acquire the benefits conferred by the money "as part of a package that may, depending on the specific circumstances, include everything from fear, horror, shame, humiliation, disgust, self-loathing ... to profound distortions of personality and of the capacity to lead a fulfilling life".

¹²⁸Scheffler 1992a, p. 381.

The first thing to note is that both of Scheffler's alleged differences are based only on psychological generalisations. So we are left with the rather odd conclusion that an agent who happened to lack the relevant feelings would be permitted to kill her own uncle. In particular, someone who had been convinced that the Hybrid View was correct might well lack such feelings.¹²⁹ So Scheffler seems forced to ground his distinction between doing and allowing in the fact that his own theory runs counter to conventional human psychology.

We should also note that, even for "psychologically normal" agents, much of the force of Scheffler's distinctions can be eliminated by changing the facts of the case. With respect to the first distinction, one can easily tell a story in which, as a result of an unforeseen change of circumstances, the only way for an agent to lay her hands on an inheritance she has always counted on receiving honestly would be for her to kill her own uncle. (For instance, perhaps the laws of inheritance have been suddenly changed.) In this case, the psychological loss involved in foregoing the inheritance would be comparable to the cost of giving up something one already possessed. With respect to the second distinction, there are many sets of circumstances which might lead an agent to feel little or no remorse at killing her uncle for profit. Many of these, of course, will be cases in which the killing was, to some degree, justified. (For instance, by the extreme cruelty of the uncle.) However, it is rather fanciful to suggest that every set of circumstances which reduces the remorse of a "normal" agent at killing a close relative is necessarily one which simultaneously provides some justification for that killing.

¹²⁹However, such a person might still possess such feelings, as strong moral feelings may co-exist with a moral view with which they are, at an intellectual level, incompatible.

We should also note that Kagan's argument does not require the total elimination of all feelings of remorse. Unless remorse at killing a close relative is always an overpowering motive for normal people, the Hybrid View will permit some such killings for personal gain. Also, remorse at killing a total stranger is likely to be much weaker than remorse at killing a close relative. So the Hybrid View will permit many killings of strangers. Neither of these conclusions seems particularly palatable.

Scheffler has a reply to this last point: "the fact that an agent had no qualms about such a killing would not suffice to show that it was really in his interest to kill, that it really advanced his good".¹³⁰

This reply is puzzling, as it seems to assume a perfectionist account of the human good, on which killing someone is bad for the agent, even if she suffers no psychological ill-effects. Now, while such an account of the human good may or may not be acceptable in its own right, it fits rather uneasily with the Hybrid View.¹³¹

To see this, we must first clarify Scheffler's claim. Scheffler may be claiming that the failure to feel qualms about killing a close relative is a sign of unhealthy psychological imbalance. This may be a reasonable empirical claim. However, it would not be sufficient to prove that a particular act of killing was bad for the agent. All that would follow from the claim would be that it was bad for the agent to be the sort of person who could kill without remorse. Given that she is such a person, she might as well gain some compensating benefit from her (on balance undesirable) condition by performing the odd remorseless killing.

¹³⁰Scheffler 1992a, p. 384.

¹³¹For recent discussions of perfectionist theories of well-being see, for instance, Griffin 1986, Hurka 1993 and Raz 1986.

So presumably Scheffler's perfectionist claim is not purely psychological. Rather, he is claiming that performing an act of killing is intrinsically bad for the agent, irrespective of any psychological effects it may have. He appears to be making a moralistic perfectionist claim. Yet it is hard to see how we can reconcile such a claim with the view, central to the original Hybrid View, that there is no general morally relevant distinction between doing and allowing. Clearly, it would be odd to say that killing is forbidden more frequently than letting die because the consequences of killing are usually worse for the agent and then to account for this difference in consequence by appealing to the intrinsic badness of killing rather than merely letting die. If killing really is intrinsically bad, then shouldn't it be forbidden for that reason alone? So Scheffler must find some way to argue that, although an act of killing is intrinsically bad for the agent, such an act is not intrinsically bad. Perhaps this can be done, but I am sceptical.

Scheffler's suggestion that killing is bad for the agent over and above its psychological costs is thus problematic. Yet, without that claim, Scheffler cannot prevent the Hybrid View from generating repugnant results for anyone with a non-standard psychology.

Even if Scheffler's line of argument enabled the Hybrid View to get the right results, it seems clear that it would get them for the wrong reasons. It seems very odd to say that the main reason why I am not permitted to kill my uncle for personal gain is because it would be bad for me! Perhaps the act of killing would be bad for me, but only because of some other feature which provided an independent reason for over-riding my agent-centred prerogative to advance my own interests. Yet, within the original Hybrid View, there can be no room for such features. On that theory, anything which cannot be explained solely in terms of impersonal value and cost to the agent cannot be explained at all.

We turn now to the last of Scheffler's alleged distinctions between doing and allowing: the different costs to others. Scheffler's presentation of this claim is explicitly collectivist: he compares the consequences of the general acceptance of killing for personal gain with the consequences of the general acceptance of letting someone die for personal gain. He sensibly suggests that the former would probably be far worse than the latter.¹³²

I discussed Collective Consequentialism at length in Chapter TWO. If Scheffler is proposing a collective analogue of the Hybrid View, then it is unlikely that such a view would be able to avoid all our previous objections to Collectivism, although a conclusive demonstration of this is difficult in the absence of a worked out Collective Hybrid View. On the other hand, if Scheffler is proposing to incorporate his collectivist reasoning into the original (individualist) Hybrid View, then he has to bridge the gap from the dire consequences of the general acceptance of killing for personal gain to the conclusion that, in the absence of such a general acceptance, this particular act of killing for personal gain will have dire consequences. I do not believe that this gap can be bridged.

I conclude, then, that Scheffler's attempt to reply to Kagan's objection within the confines of the original Hybrid View does not succeed.

5.2.3: An Alternative Response to Kagan

In his 1992 article, Scheffler discusses another, more radical, response to Kagan's objection: the adoption of a "modified" Hybrid View.¹³³ The modification consists of replacing the original "pure-cost" agent-centred prerogative with a "no-harm" agent-centred prerogative, which would allow the agent to give disproportionate weight to her own

¹³²Scheffler 1992a, pp. 384-385..

¹³³See Scheffler 1992a, pp. 387-395.

interests only when considering an act which did not directly harm anyone else. Harming others for personal gain would thus be ruled out, but agents would still be permitted to harm others in pursuit of impersonal value. As Scheffler does not present his Modified Hybrid View in any great detail, a comprehensive evaluation would perhaps be premature. I shall confine myself to a few brief observations.

We have been regarding the rejection of the distinction between doing and allowing as a central feature of Consequentialism. The Modified Hybrid View contains a partial form of this distinction. It thus sits uneasily between Consequentialism and non-Consequentialism. The first question to address would be whether or not the Modified Hybrid View represents, for our purposes, an unacceptably great departure from Consequentialism. We are interested in responses to the Demandingness Objection from within Consequentialism. If the latest modifications take the Hybrid View beyond Consequentialism, then we should look elsewhere.

A related problem for the Modified Hybrid View is to find a plausible rationale for incorporating a no-harm prerogative which does not also commit us to the full distinction between doing and allowing. That is, can we justify forbidding all harmings for personal gain while allowing all harmings in pursuit of impersonal value? As Scheffler himself notes "it is hard to see how there could be considerations that supported this claim without implying that harming is in general worse than allowing harm".¹³⁴

Scheffler does not claim to provide such a rationale. However, he does seek to show that one might be available. He discusses two possible candidates. The first arises from the "adjusted consequential considerations" discussed in the previous section. If acts of harming for personal gain really are more costly (both for the agent and for others) than

¹³⁴Scheffler 1992a, p. 388.

acts of allowing someone to be harmed for personal gain, then this fact may justify a no-harm prerogative.

Unfortunately, our discussion of these considerations above suggests that they will be unable to do the work which Scheffler proposes for them. In particular, it is unlikely that we could establish the claim that all acts of harming for personal gain are more costly (either for the agent or for others) than acts of allowing someone to be harmed for personal gain, without illicitly appealing to the assumption that harming is undesirable. Yet, once this assumption is made, these considerations will provide as much support for a ban on harming in pursuit of impersonal value as they do for a no-harm prerogative.

Scheffler's second suggestion is to appeal to the "quasi-practical advantages that prohibitions against harming have in comparison with requirements to prevent harm".¹³⁵ For instance, the former are easier to teach, to internalise and to obey than the latter. This suggestion has obvious affinities with Scheffler's earlier appeal to the collective costs of a general acceptance of harming for personal gain. Once again, we will be faced by all the difficulties we discussed at length in Chapter TWO. Furthermore, it is not clear that quasi-practical advantages could ever be sufficient to generate a rationale for a no-harm prerogative. In particular, Scheffler's approach here seems too question-beggingly piecemeal. He assumes that we already have in place a general requirement to promote the good, and the proviso that departures from that requirement are only to be permitted (and never required) by appeals to the agent's own interests. We then seek to draw a line between permissible and impermissible types of non-optimal, self-interested behaviour. It is only at this point that quasi-practical considerations are allowed to play their part. Now, either these considerations are strong enough to provide a compelling rationale for the no-harm prerogative or they are not. If they are not, then Scheffler will have failed to

¹³⁵Scheffler 1992a, p. 389.

demonstrate that such a rationale may be possible. On the other hand, if the quasi-practical considerations can provide a rationale for the no-harm prerogative, then why shouldn't they be brought into play earlier in the construction of our moral theory? Yet, if these considerations were brought into play earlier, and if they are so strong, then how can we be sure that they (or other similar considerations) won't also provide a rationale for some constraint on the performance of harm in pursuit of impersonal value? If we cannot be sure that this won't happen, then we are not entitled to conclude, as Scheffler does, that the possible strength of the consideration establishes the possible existence of a rationale for the no-harm prerogative.

I have not established the impossibility of providing a rationale for the no-harm prerogative. However, it is not clear that Scheffler has established the possibility of doing so either.

From our perspective, a more general problem with the Modified Hybrid View is that it seems set to increase the demands faced by agents, rather than reducing them. After all, the only difference between the original Hybrid View and the modified version is that the latter, in addition to forbidding some sub-optimal allowings of harm, forbids all sub-optimal harmings.

Proponents of the Modified Hybrid View may reply that the introduction of a constraint on sub-optimal harmings will allow us to make our prerogative more generous in non-harming situations. For instance, we can now permit Kagan's inheritor to allow a distant stranger to die rather than sacrificing \$10,000, safe in the knowledge that this does not commit us to allowing her to kill her uncle in order to inherit a similar sum.

There is certainly something in this reply. The Modified Hybrid View may be able to exchange a more demanding approach to harming for a less demanding approach to allowing harm. However, we should note that, unless our defence of the Modified Hybrid View is much stronger than our previous discussion suggests, it is unlikely that we will be able to establish a total ban on harming for personal gain. The question of precisely how much harming is permitted will thus depend in part on the stringency of our general prerogative. If we are not to admit too much harming, we cannot afford to be too generous with respect to the allowance of harm.¹³⁶

It is too soon to rule the Modified Hybrid View out completely. However, our discussion has cast serious doubt on its claim to provide an adequate Consequentialist solution to the Demandingness Objection.

The only remaining option for Scheffler would be to introduce deontological constraints such as those advocated by Murphy. However, in Scheffler's vocabulary this would amount to the introduction of agent-centred restrictions. It would thus be an abandonment of the Schefflerian project.

Strictly speaking, this would be an abandonment of our project also. However, we should note that such a constraint is unlikely to be absolutely deontological. Unless we take an extreme view, we will not end up with a ban on killing per se, irrespective of consequences. Rather, our constraint will be on killings with specified consequences, where the same consequences would be sufficient to justify an agent's letting someone die.

¹³⁶We should also note in passing that the Modified Hybrid View seems no better equipped than Scheffler's original Hybrid View to cope with objections such as the Wrong Facts Objection.

Perhaps we could construct a plausible semi-Consequentialist alternative to Scheffler's position, one which combined some instances of the Scheffler Asymmetry with some non-absolute constraints. Such a view would not be as far removed from the original Hybrid View as would a fully deontological theory. The introduction of some partial constraints, while representing a departure from Consequentialism, might be compatible with the spirit of our Consequentialist project, providing the constraints did not become too pervasive.¹³⁷ A full discussion of the possible roles of such limited constraints within a Consequentialist theory is beyond the scope of this thesis. Accordingly, I shall not pursue this possibility any further at this stage.

I conclude that Scheffler has not demonstrated that the Hybrid View is able to respond adequately to Kagan's objection. Unless we introduce some formal constraints, any moral theory which contains an agent-centred prerogative must sometimes permit agents to kill others (at the expense of the general good) in order to further their own ends. Such permissions do not seem to be acceptable.

5.3: Conclusions

I conclude that the Hybrid View is not acceptable as it stands. In particular, it is unable to provide an adequate response to the Demandingness Objection. However, the Hybrid View does seem to come closer to satisfying our intuitions than any of the other theories we've discussed in Part ONE. For that reason, we will discuss a revised version of the Hybrid View in Chapter SEVEN.

¹³⁷In particular, if the constraints required us to introduce a general distinction between doing and allowing, then this in turn might undermine our Consequentialist response to the Demandingness Objection. As I argued in my discussion of Scheffler's Modified Hybrid View, such a move threatens to undermine the Consequentialist basis of the Hybrid View.

Part TWO: Towards a New Solution

Chapter SIX: Non-Proportionality Introduced

In Part TWO, I sketch some new responses to the Demandingness Objection. The basis for these is the theoretical device of non-proportionality, which is introduced in this chapter.

6.1: Non-Proportionality Introduced

The theoretical innovation discussed in this chapter is a distinction between proportional and non-proportional accounts of wrongness. A central feature of Consequentialism is the claim that the wrongness of acts depends solely on the value of outcomes. We shall now focus more closely on the relationship between wrongness and values. That relationship can be broken down into three components, which take us from the values of outcomes to the wrongness of acts. The first component is the "beneficence" of an act, a measure of the amount of good it produces, which is a function of the value (or expected value) of the outcome of performing that action. We then generate the agent's "promotion-of-good" reason for performing that act rather than some other act available in that situation. This will be a function of the different levels of beneficence of the two acts concerned. Finally, under any Consequentialist theory, the wrongness of an act will be a function of the agent's promotion-of-good reasons in that situation.¹³⁸

Each of our three functions can take a number of possible forms. I shall choose standard versions of the first two functions, in order to focus on the third.¹³⁹ Assume that A (an

¹³⁸Any moral theory which makes some room for promotion-of-good reasons can be said to be partially Consequentialist. A fully Consequentialist theory will be one which goes on to claim that all moral reasons are promotion-of-good reasons.

¹³⁹These standard functions are implicit in most extant versions of Consequentialism, though I admit that they are rarely made explicit.

agent) faces a choice between three acts: x (which will lead to outcome One), y (which will lead to outcome Two), and z (which will lead to outcome Three). The values of One, Two and Three are 50, 100 and 200 respectively. On our standard account of the first function, the beneficence of each act will be directly proportional to the value of its outcome. For instance, we might choose a function on which the beneficences of x, y and z are 50, 100 and 200 respectively. A's promotion-of-good reason to do one act rather than another is then directly proportional to the difference in beneficence between the two acts. For instance, A's reason to do y rather than x will have a strength of 50, whereas her reason to do z rather than x will have a strength of 150. The important factor is the ratio between the strengths of the various reasons. In this case, A's reason to do z rather than x is three times as strong as her reason to do y rather than x, because there is three times as much value at stake in the former decision as in the latter.

We turn now to the relationship between reasons and wrongness. In any Consequentialist theory, this function is governed by the following rule:

The Constraint : the wrongness of an act in a given situation is a function only of the agent's various promotion-of-good reasons in that situation.

This basic constraint can be fleshed-out in a variety of ways. Two plausible alternative constraints are as follows:

The Proportionality Constraint : for any two actions (x and y) available to an agent (A), the wrongness of A's doing x rather than y will be directly proportional to the strength of A's promotion-of-good reason to do y rather than x (and, hence, directly proportional to the difference in beneficence between y and x).

and:

The Ordinality Constraint : Let x , y and z be any three acts available to A , where z is the most beneficent. It will then be more wrong for A to do x than for A to do y if, and only if, her reason to do z rather than x is stronger than her reason to do z rather than y (and, hence, if and only if y is more beneficent than x).

Clearly, any account of wrongness which satisfies the Proportionality Constraint will satisfy the Ordinality Constraint. However, the converse is not the case. An account of wrongness can satisfy the Ordinality Constraint without satisfying the Proportionality Constraint.¹⁴⁰ The Proportionality Constraint is a stronger requirement.

Traditional Individual Consequentialism usually implicitly adopts the Proportionality Constraint. The wrongness of our actions is thought to be directly proportional to the amount of good at stake. However, Consequentialism itself requires only the Ordinality Constraint. Indeed, I shall argue that replacing the Proportionality Constraint with the Ordinality Constraint may enable Consequentialists to solve a number of pressing problems, including the Demandingness Objection.

The Ordinality Constraint leaves open the precise relationship between the wrongness of acts and the strengths of reasons. We can thus combine the Ordinality Constraint with a more precise thesis regarding that relationship. The most interesting contenders are the following:

¹⁴⁰To see this, let w be the degree of wrongness of A 's doing x , and let s be the strength of A 's promotion-of-good reason to do z rather than x , where z is the most beneficent act available to A in that situation. Both theses say that w is some function (f) of s . Under the Ordinality Constraint, f is a strictly increasing function. Under the Proportionality Constraint, f is a linear function with positive gradient. Now, if f is linear with a positive gradient, then f will be strictly increasing. Therefore the Proportionality Constraint implies the Ordinality Constraint. However, f can be strictly increasing without being linear. Therefore, the Ordinality Constraint does not imply the Proportionality Constraint.

The Proportionality Thesis : this replaces the Ordinality Constraint with the Proportionality Constraint. The wrongness of an action is a strictly increasing linear function of the strength of the agent's reason to perform the most beneficent act instead. The Proportionality Thesis is part of traditional Consequentialism.

The Undefined Thesis : the ordinal ranking required to satisfy the Ordinality Constraint is the only measure of the wrongness of an action which can coherently be defined. We can say that one act is more wrong than another, but we cannot say by how much.

The Irrelevance Thesis : the ordinal ranking required to satisfy the Ordinality Constraint is the only measure of the wrongness of an action which has any moral significance. We only need to be able to say that a given act is more wrong than another. We might be able to say how much more wrong it is, but such a judgement would be morally irrelevant.

The Decreasing Thesis : the wrongness of an action increases as the strength of the agent's reason to perform the most beneficent act instead increases. However, the size of the increase in wrongness decreases steadily with each successive increase in the strength of that reason.

The Increasing Thesis : the wrongness of an action increases as the strength of the agent's reason to perform the most beneficent act instead increases. Moreover, the size of the increase in wrongness increases steadily with each successive increase in the strength of that reason.

I shall be focusing on the contrast between the Proportionality Thesis and the Decreasing Thesis. First, however, we must briefly discuss the Undefined Thesis and the Irrelevance Thesis.

In the absence of some argument as to why differences between degrees of wrongness cannot be defined, the Undefined Thesis is arbitrary. I cannot think of any convincing argument for that view. Therefore, I propose to put the Undefined Thesis to one side.

The following argument can be advanced in favour of the Irrelevance Thesis. Assume that I face a meta-choice between the following two choices (i.e., I can choose which choice I will be faced with):

	Option	Value
Choice One	x	100
	y	90
Choice Two	w	100
	z	50

Initially, it may seem that the two choices differ. On the hand, if I choose x in Choice One, I will feel secure in the knowledge that, even if I'd chosen y, I'd only have been slightly worse-off. On the other hand, if I choose w in Choice Two, I may feel satisfied that, by choosing correctly, I have increased my utility by 50, rather than merely increasing it by 10, as would have been the case in Choice One.

However, assume that any values attached to such feelings have been included in the values attached to the outcomes. It seems to follow that, if I am rational, I should be indifferent between the two choices. If my meta-choice is for Choice One, then I will go on to choose option x. If my meta-choice is for Choice Two, then I will go on to choose option w. Either way, my expected utility is 100. The fact that the difference between w and z is far greater than the difference between x and y is irrelevant. The only comparison is between x and w. The opportunity cost of the superior option (i.e., the amount of good foregone in choosing that option rather than its competitor) is immaterial.

To illustrate this, consider the following two meta-choices:

Meta-choice Two:	Option	Value
Choice One	x2	101
	y2	91
Choice Two	w2	100
	z2	50

Meta-choice Three:	Option	Value
Choice One	x3	100
	y3	90
Choice Two	w3	101
	z3	51

In each of these meta-choices, the opportunity cost of each option is the same as in the original meta-choice. However, it seems clear that, in Meta-choice Two, I should opt for Choice One, as I can then choose x2 for a utility of 101, rather than the utility of 100 which is the best I could hope for in Choice Two. By contrast, in Meta-choice Three, I should opt for Choice Two, as I can then choose w3 for a utility of 101 rather than the utility of 100 which is the best I could hope for in Choice One. The slightest difference between the values of the superior option in each choice is sufficient to unbalance the meta-choice, whereas no difference in opportunity cost has that effect.¹⁴¹

The example lends support to the Irrelevance Thesis, as it suggests that all that matters is which option will produce the most good. The opportunity cost of that option (i.e., the amount of good the next best option would have produced) is morally irrelevant.

However, there are other cases where the amount of good foregone does seem to be morally relevant. Assume that I face each of the choices from my first meta-choice in order:

	Option	Value
Choice One	x	100
	y	90
Choice Two	w	100
	z	50

¹⁴¹This example has been expressed in terms of individual rationality. However, the same would hold true of morality. To see this, simply replace the talk of personal utility with talk of impartial moral values.

Suppose that in Choice One I do *y* while in Choice Two I do *z*. One each occasion, I fail to produce the best possible outcome. Under the Irrelevance Thesis, this is the only relevant fact. Therefore, my response to Choice One is just as wrong (if it is wrong at all) as my response to Choice Two. This seems incorrect. Surely there is an important sense in which my second choice is more wrong than my first, precisely because I forego more good on that occasion.

We must clarify the notion of "amount of good foregone". This does not mean the amount of good which would have been produced by the best action of those not chosen. Rather, it relates to the difference between that amount of good and the amount of good which is actually produced. In our present example, the first quantity is the same for both choices (100), but the second is greater in Choice Two (50) than in Choice One (10). My suggestion is that the wrongness of an act is in some way dependent upon the amount of good foregone, in this second sense.

We can capture this suggestion in terms of differences between strengths of reasons. My reason to do *x* has the same strength as my reason to do *w*. However, my reason to do *y* is stronger than my reason to do *z*. The difference in strength between my two reasons in Choice One is thus less than the difference in strength between my two reasons in Choice Two. If we say that wrongness increases as the difference in reasons increases, we can say that I act more wrongly in Choice Two than in Choice One.

The Irrelevance Thesis is certainly more plausible than the Undefined Thesis, and it does highlight some important points, in particular the following:

- (i) When faced with the first meta-choice, there is no (promotion-of-good) reason to choose Choice One over Choice Two; and

(ii) The beneficence of x in Choice One is identical to the beneficence of w in Choice Two.

However, the Irrelevance Thesis fares less well with comparisons of reasons. In particular, we have just seen that it has difficulty with the following (plausible) claim:

(iii) My reason to choose w *over* z in Choice Two is stronger than my reason to choose x *over* y in Choice One.

We therefore want a theory which retains (i) and (ii) while accommodating (iii). Both the Decreasing Thesis and the Increasing Thesis would provide such a theory. I shall not discuss the latter in any detail here, as I think it has little or no intuitive appeal, particularly as it would only serve to make Consequentialism even more demanding! Instead, I shall focus on the Decreasing Thesis.¹⁴²

We can define the Decreasing Thesis as the conjunction of the following claims (where x , y , and z are three options facing an agent A , and where z is the most beneficent of those options):

The Scalar Condition : Wrongness is a scalar notion. Not only are some acts wrong while others are not wrong, but different wrong acts will have different degrees of wrongness. Some acts will be more wrong than others.¹⁴³

¹⁴²However, I shall return to the Increasing Thesis in Chapter SEVEN (section 7.3) where I discuss an analogous thesis in relation to the appeal to cost.

¹⁴³A similar notion is developed in Slote 1985, pp. 76-91.

The Ordinality Condition : x will be more wrong than y if, and only if, A's reason to do z rather than x is stronger than A's reason to do z rather than y.¹⁴⁴

The Decreasing Condition : If x is more wrong than y, then the ratio of the wrongness of x to that of y will be GREATER than the ratio of the strength of A's reason to do z rather than x to the strength of her reason to do z rather than y.¹⁴⁵

The Ordinality Condition distinguishes the Decreasing Thesis from the Irrelevance Thesis, while the Decreasing Condition distinguishes the Decreasing Thesis from the Proportionality Thesis.

The Decreasing Thesis can be quite confusing, so let's illustrate it using a simple example. Assume that we have a single agent (A) who faces a one-off choice between three options (x, y and z), where the values and beneficences of the options, and the strengths of A's reasons, are set out in the table below:

Act	Value	Beneficence	Reason to do z instead
z	100	100	0
y	75	75	25
x	50	50	50

¹⁴⁴In what follows, it is sometimes helpful to speak of the wrongness of the most beneficent act, in this case z. If we adopt the convention that A's reason to do z rather than z has a strength of zero, then we can also say that z has a zero degree of wrongness.

¹⁴⁵Conditions governing comparisons of wrongness between acts available to different agents, or between acts available to the same agent in different situations at different times, can also be constructed. For simplicity's sake, I have focused primarily on the single agent, one-off situation case.

The Ordinality Condition tells us that x is more wrong than y , which is more wrong than z , due to the relative strengths of A 's reasons. The Decreasing Condition tells us that the ratio of the wrongness of x to that of y will be greater than the ratio of the strength of A 's reason to do z rather than x to the strength of her reason to do z rather than y . The latter ratio is 2:1. Therefore, x will be more than twice as wrong as y .

Under the Proportionality Thesis, by contrast, the two ratios would be identical, as wrongness is directly proportional to the strength of the agent's reasons. Therefore, x would have been exactly twice as wrong as y . x is thus more wrong, compared to y , under the Decreasing Thesis than under the Proportionality Thesis. The Decreasing Thesis is also less precise than the Proportionality Thesis, as it doesn't tell us exactly how much more wrong one action is than another.

6.2: Justifying the Decreasing Thesis

So far I have only explained what the Decreasing Thesis is. I have not provided an argument in support of it. In this chapter, I am more interested in exploring the Decreasing Thesis than in justifying it. However, in this section I hope to present some considerations in favour of replacing the Proportionality Thesis with the Decreasing Thesis.

The first consideration is that, although Consequentialists have always implicitly adopted the Proportionality Thesis, they have often done so for bad reasons. One such reason is that Consequentialists may have confused the Proportionality Thesis with the following:

The Proportionality of Value Thesis : Let A and B be any two outcomes. If B contains ten times as much of whatever it is that makes an outcome valuable as A , then the value of B will be ten times the value of A . For instance, assume that what makes an outcome valuable

is the existence of human lives which are worth living. Let A be an outcome in which there are 40 million people have lives which each have a value of 50. Let B be another outcome in which the same people all have lives with a value of 100. Under the Proportionality of Value Thesis, the value of outcome B will thus be twice the value of outcome A, as B contains twice as much of whatever it is that makes lives worth living as A.

I shall not discuss the plausibility of the Proportionality of Value Thesis here. I wish only to point out that Consequentialists who adopt this thesis need not adopt our original Proportionality Thesis. By separating the various component functions which take us from assessments of value to assessments of wrongness, supporters of the Decreasing Thesis can deny that an action which fails to bring about outcome B is twice as wrong as one which fails to bring about outcome A, without being forced to conclude that the value of B is less than twice the value of A.¹⁴⁶ Indeed, this is the whole point of the Decreasing Thesis. Consequentialists' attachment to the Proportionality of Value Thesis is thus no barrier to the rejection of the Proportionality Thesis. The Decreasing Thesis need not imply eccentric views about the values of outcomes.

Another reason why traditional Consequentialists ignore the Decreasing Thesis is that it seems to be no different, in practice, to the Proportionality Thesis. Consequentialists are concerned primarily with the question of whether one act is more or less wrong than another. Yet any theory which meets the Ordinality Constraint will give the same answer to that question: of any two acts, the one with the lower beneficence is the more wrong. This will be true whether we adopt the Proportionality Thesis or the Decreasing Thesis.

However, I shall demonstrate below (in section 6.3) that the shift from the Decreasing

¹⁴⁶ Assuming that we can tell some not too fanciful story in which agents are choosing between one of A and B and some option with a value of zero.

Thesis to the Proportionality Thesis can be very significant, especially within a moral theory which offers scalar accounts of key moral terms.

Many will reply that, even if Consequentialists have adopted the Proportionality Thesis for the wrong reasons, the Proportionality Thesis is still the overwhelmingly more plausible interpretation of the relationship between values and wrongness.

I am tempted to reply that it is not clear why this should be so. Why is a linear function from strengths of reasons to degrees of wrongness intrinsically more plausible than a non-linear one? Defenders of the Proportionality Thesis certainly cannot simply rest content with their intuitions. A positive argument for the Proportionality Thesis is required.

However, this reply may seem unconvincing. On the face of it, the Proportionality Thesis simply is a far more natural interpretation than the Decreasing Thesis. The need for an argument for the Decreasing Thesis is thus much more pressing than the need for an argument for the Proportionality Thesis.

My principal argument for the Decreasing Thesis is the fact that it enables Consequentialists to provide plausible responses to the Demandingness Objection. If a Consequentialist theory incorporating the Decreasing Thesis provides a better explanation of our reflective moral intuitions than any Consequentialist theory which incorporates the Proportionality Thesis, then Consequentialists should replace the Proportionality Thesis with the Decreasing Thesis. A complete comparison of the theoretical virtues of the two theses is beyond the scope of this thesis. However, one strong moral intuition is that morality should not be as demanding as traditional Consequentialist theories (which all incorporate the Proportionality Thesis) seem to make it. In the remainder of this thesis, I seek to show that the Decreasing Thesis provides a significantly less demanding version of

Consequentialism than any extant theory based on the Proportionality Thesis. If I am right, then this would provide a strong argument in favour of the Decreasing Thesis.

Even if the Decreasing Thesis cannot provide a complete solution to the Demandingness Objection, it may provide part of a solution. In Chapter SEVEN, I show how the introduction of non-proportionality might improve Scheffler's Hybrid View, while Chapter EIGHT explores possible non-proportional versions of Parfit's account of blameless wrongdoing. If we are unconvinced by the use to which non-proportionality is put here, we may view the present chapter as a necessary preliminary to those later discussions.

One final consideration is that, when a theory is as prominent as Consequentialism, it is a worthwhile exercise to examine all its theoretical resources, especially those which have been hitherto neglected, even if we cannot as yet provide a conclusive argument why those particular resources should be used. At the very least, a thorough examination of an option we may not end up endorsing will deepen our understanding of those options we do endorse.¹⁴⁷

I am confident that, taken together, these various considerations provide sufficient reason for considering the Decreasing Thesis in some detail, as I now propose to do.

¹⁴⁷We should note that our theory is developed in three separate stages. An account of beneficence, an account of reasons, and finally an account of wrongness. What will do the work in defeating the Demandingness Objection is the inclusion of a non-proportional element in the account of wrongness. I have introduced that element in the account of wrongness. However, we could equally well have introduced an element of non-proportionality at either of the two earlier stages, giving us either a decreasing function from the values of outcomes to the beneficence of acts, or a decreasing function from comparisons of beneficence to the strengths of promotion-of-good reasons. For the purposes of our discussion of wrongness, the end result would be much the same, but the novel account of reasons, in particular, might yield additional complications.

6.3: Applying the Account to Affluent

We now ask how this new account would deal with Affluent's situation. We may not be able to tell precisely how large a donation to Oxfam would produce the best possible results in the long run, all things considered. However, we do know that the best thing for Affluent to do would be to donate much more than she has done, or else the Demandingness Objection would not arise. It is clear, then, that Affluent does not do what is most beneficent. Similarly, she doesn't do what she has most reason to do, as that would be whatever is most beneficent.

We now consider whether or not Affluent has done wrong. To simplify our discussion, assume that Affluent has three options:

Best : make the optimal donation. 10 lives are saved.

Second-Best : make a donation of 10% of her income, keep the rest for herself. One life is saved. Affluent has a good time.

Third-Best : make no donation. No lives saved. Affluent has a better time.

Assume, further, that Affluent actually opts for Second-Best. If Affluent had chosen Best, then she wouldn't have done anything wrong. If she chooses either Second-Best or Third-Best, she does something wrong. As Affluent actually does Second-Best, then we must conclude that she's done wrong. The question is: how wrong?¹⁴⁸

¹⁴⁸We should note that, insofar as it is a separate question, the issue of whether or not Affluent is to be blamed for her actions is discussed in detail in Chapter EIGHT.

It is at this point that the two unconventional elements in our account of wrongness appear. Because wrongness is scalar, we know that donating no money at all would have been more wrong than donating some. We can therefore distinguish Affluent's behaviour from that of someone who donates nothing at all. Furthermore, our non-Proportional element says that the wrongness of a given option depends upon the difference between the beneficence of that option and the beneficence of the best available act. It therefore depends upon the difference between the value of the outcome of performing the best act and the outcome of performing the option concerned. However, differences in wrongness are not proportional to differences in beneficence, nor to differences in the strengths of reasons. Donating no money (Third-Best) may thus be much more wrong than donating some (Second-Best).

To illustrate this point, let us oversimplify our story still further and focus our attention on the number of lives saved from starvation in each possible outcome. If Affluent donates all her money, ten lives are saved. So Best produces an outcome with a value of ten. If Affluent donates some of her money, one life is saved. So Second-Best produces an outcome with a value of one. If Affluent donates no money, no lives are saved. So Third-Best produces an outcome with a value of zero.

The difference between the beneficence of Best and that of Second-Best is thus nine units, whereas the difference between the beneficence of Second-best and that of Third-Best is only one unit. On a proportional account, Second-Best would be ninety-percent as wrong as Third-Best. The difference in wrongness between the two imperfect options would be much less than the difference between either and the best option. However, on a decreasing account of wrongness, Second-Best will be less than ninety-percent as wrong as Third-Best. Indeed, if our non-proportional element is strong, then Second-Best may be much less than ninety-percent as wrong as Third-Best. The two options need not be very close

together in terms of wrongness, even though the values of their outcomes are much more similar to each other than either is to the value of the outcome of the most beneficent act.

Our scalar and non-proportional elements thus enable us to judge Affluent much less harshly than do traditional Consequentialist accounts, as we can say, not only that donating no money at all would have been more wrong than what she does, but that it would have been much more wrong.

6.4: Conclusions

I conclude that the shift from proportionality to non-proportionality provides at least a partial response to the Demandingness Objection. The question of whether or not this response is sufficient is explored in the next chapter.

Chapter SEVEN: a non-proportional Scheffler

Some will not be satisfied with the response sketched in Chapter SIX, as the Demandingness Objection has not been completely dissolved. This may suggest that my approach is on the wrong track. We can see this by comparing my solution with Slote's. Recall that, in our discussion of Slote's Satisficing Consequentialism in Chapter FOUR, we concluded that cost to the agent, not the amount of good she had previously done, was the relevant factor in determining how much was expected of her. The amount of good already done was relevant only as a rough approximation of cost.

It seems that we can present an analogous argument against my interim solution. Consider the following version of our Achilles' Tale:¹⁴⁹

The Two Heroes' Tale : two heroes (Achilles and Patroclus) are placed in separate rooms. Achilles has three choices, press button A and thereby save no one from starvation, press button B and thereby save 400 people from starvation, or press button C and thereby save 401 people from starvation. None of these options will cost Achilles anything. Patroclus has two choices. Press button D and thereby save no one from starvation, or press button E and thereby save 1 person from starvation. Pressing button D will cost Patroclus nothing, but pressing button E involves a significant cost to him.

Assume that Achilles presses button B, saving 400 people rather than 401, and that Patroclus presses button D, saving no one rather than one person. It seems that, on my decreasing account of wrongness, what Achilles does is less wrong than what Patroclus does. This is because, as the amount of value at stake increases, the wrongness involved in failing to produce a given unit of additional value decreases. Therefore, saving 400 people

¹⁴⁹See section 4.3 of Chapter FOUR.

from starvation when you could have saved 401 is less wrong than saving nobody when you could have saved one person from starvation. Furthermore, the cost to the agent plays no role in this theory. So the fact that Achilles could have saved his extra person at no extra cost to himself, whereas Patroclus would have faced a significant personal cost, is irrelevant.

It may seem that such a theory gets things the wrong way round. The most relevant factor in this situation is surely that Achilles could have saved his extra person at no cost whatsoever to himself, whereas Patroclus could only have saved his extra person at significant personal cost. What Achilles does is thus much more wrong than what Patroclus does. It seems, therefore, that my theory is seriously defective.

There are a number of possible responses to this objection. In the first place, we should note that, unlike Slote, I do not say that Achilles has no reason to save his additional person. If he fails to save that person, Achilles will have done something wrong. So my solution is not as defective as Slote's.

A second response would be to deny that appeals to cost carry any normative force. This is the line taken by Shelly Kagan, and followed up by Roger Crisp.¹⁵⁰ However, our discussion of Slote's Satisficing Consequentialism in Chapter FOUR suggests that considerations of cost almost certainly are relevant to moral judgements. Furthermore, the Demandingness Objection itself is premised on the idea that considerations of cost should carry moral weight. If we take that objection seriously, then at the very least we should explore the option of a cost-sensitive non-proportional theory.

¹⁵⁰See Kagan 1989, pp. 231-270, and Crisp 1994.

A final response to the cost objection would be to supplement my approach with a more cost-sensitive element. In this chapter, my primary purpose is to discuss one such supplementation: a non-proportional version of Scheffler's agent-centred prerogative.

7.1: Developing a Non-proportional Hybrid View

Scheffler's original Hybrid View is an attempt to supplement traditional Consequentialism with a recognition of the moral significance of the costs to the agent. In Chapter FIVE, I argued that the original Hybrid View is unacceptable. However, it is worth asking whether or not a hybrid version of our non-proportional Consequentialism would fare any better.

Under our original interpretation of the agent-centred prerogative, the maximum sacrifice required from an agent was proportional to the marginal good which would be produced by donating an additional dollar to charity. A tenfold increase in the amount of good which a given donation would produce thus required the sacrifice of interests which were ten times as important to the agent. This suggests an analogue of the Proportionality Thesis. By contrast, on a non-proportional agent-centred prerogative the maximum sacrifice required of the agent would increase as the amount of good at stake increased. However, the former might not increase as rapidly as, or in proportion to, the latter. Another additional feature of our account will be a scalar element; some acts will be more wrong than others, in line with the account of wrongness developed in Chapter SIX.

There are two questions to ask concerning this revised Hybrid View:

(1) What will a non-proportional version of Scheffler's agent-centred prerogative look like?

(2) Would the resulting hybrid provide an adequate response to the Demandingness Objection?

If we are interested in the relationship between our existing account of wrongness and the appeal to cost, then we have two different routes to explore: tinkering with the appeal to cost or tinkering with that account. Let us begin with the first: a non-proportional version of the appeal to cost. In particular, I shall focus on a revision of the Hybrid View which combines an increasing account of the appeal to cost with a proportional account of the other components of wrongness.

To develop this new theory, let M be the disproportionate weighting the agent is allowed to give to her own interests under the Hybrid View. Under Scheffler's original Hybrid View M is a constant. The amount of extra weight the agent is allowed to give to her own interests is the same regardless of the cost involved. Under our new theory, by contrast, M is a variable factor determined by an increasing function of the cost to the agent. As the cost the agent would have to bear to produce a given additional amount of good increases, the amount of weight she's allowed to give to her own interests also increases.

To illustrate this view, assume that A faces a choice between three options (x , y and z). In the table below, the first column represents the value of the outcome of each action. The second column represents the cost A would have to bear if she performed that action instead of performing z , the least costly option. The third column tells us how much weight A is allowed to give to her own interests at that level of cost. The fourth column gives us the total weight of the cost the agent would bear if she performed that option (adding each successive unit multiplied by its weight).¹⁵¹ The fifth column tells us whether or not A is

¹⁵¹For this example, I make the simplifying assumption that each of the first 10 units of cost has a weighting of 2, while each subsequent unit has a weighting of 5. In a more realistic theory, this simple

allowed to perform z instead of the action in question (by subtracting the weighted cost from the amount of value at stake).

Act	Value	Cost	Weight	Total Weight	Permitted to do z?
x	100	17	5	55	YES
y	75	10	2	20	NO
z	50	0	0	0	-

Under a non-proportional version of the Hybrid View, the weighting at a cost of 10 must be the same as at a cost of 17. Therefore, if A is not allowed to perform z instead of y, then she won't be allowed to perform z instead of x, as the extra cost is less, while the amount of additional value at stake is the same. By contrast, under our new non-proportional theory, A is allowed to avoid x, even though she's not allowed to avoid y, as the difference in cost between y and x counts for more than that between y and z.

This combination of moral judgements would be impossible under any traditional Hybrid View. I shall argue below that this facility enables the new theory to defeat the Demandingness Objection.

Let us turn now to the second version of the non-proportional agent-centred prerogative. This combines a traditional proportional account of the appeal to cost (such as that discussed in Chapter FIVE) with the non-proportional account of wrongness outlined in Chapter SIX. On this view, wrongness will be judged by weighing two factors: the weighted value (as measured by an increasing, but non-proportional, function of the value

stepwise function would be replaced by a more continuous one, but this does not affect the argument of the text.

of the outcome of the act) versus the weighted cost (as found by multiplying the cost to the agent by a constant factor, as in Scheffler's original Hybrid View).

The agent will be allowed to perform a given action (y) rather than another action (x) if, and only if, the weighted cost of doing x rather than y exceeds the weighted value of doing x rather than y. As the amount of value at stake increases, the amount of cost the agent is required to bear increases also. However, the latter increases more slowly than the former. For instance, if the amount of value at stake increases tenfold, then A will be required to bear a greater cost, but one which is less than ten times greater, than she would initially have been required to bear. To illustrate this, return to our previous example, where A faces a choice between x, y and z. As before, the first two columns represent the value and cost of each function. The third column represents the disproportionate weighting A is allowed for each unit of cost (this time, this is the constant M). The next column gives the weighted cost (found simply by multiplying the cost by M). The fifth column gives us the weighted value (a decreasing function of the value of the outcome). The sixth column tells us whether or not A is allowed to do z instead of the action in question (this time found by comparing the weighted cost in column four with the weighted value in column five).

Act	Val	Cost	M	W'ed Cost	W'ed V	Permitted to do z?
x	100	17	2	34	30	YES
y	75	10	2	20	25	NO
z	50	0	2	0	0	-

The judgements represented in the last column are the same as those obtained in the first version of our non-proportional Hybrid View. Once again, these would not have been possible under any proportional account of the Hybrid View.

We must now choose between these two possible non-proportional agent-centred prerogatives. Let a value-based theory be any moral theory which combines a non-proportional account of the reason to promote the good with a proportional account of the appeal to cost. Let a cost-based theory be any moral theory which combines a proportional account of the reason to promote the good with a non-proportional account of the appeal to cost. We begin by focusing our attention on cases where the two accounts diverge. Such cases will be of two types, which are illustrated by the following two stories:

The Same Values Story : Two agents (A and B) each face a choice between two different options (x and y and w and z respectively). The values and costs involved are set out in the following table:

Agent	Act	Value	Cost
A	x	100	20
	y	50	0
B	w	100	10
	z	50	0

The Same Costs Story : essentially the same, except that the values and costs involved differ as set out in the following table:

Agent	Act	Value	Cost
A	x	100	10
	y	50	0
B	w	150	10
	z	50	0

We begin with the Same Values Story. Here, the amounts of value at stake are exactly the same in A's case as in B's. So any differences in wrongness must be attributable to differences between the costs involved. On a value-based theory, the appeal to cost is a traditional proportional one. Therefore, in a case where the values are unaltered, wrongness must be proportional to cost. If B does z rather than w and A does y rather than x, then B will be doing something twice as wrong as what A does. Any value-based theory must deliver the same moral judgements in the Same Values Story as a fully proportional theory.

In a cost-based theory, by contrast, wrongness is not proportional to cost, even when values are unaltered. Therefore, if B does z and A does y, then B will do something which is more than twice as wrong as what A does, as the second ten units of cost will count for more than the first, so A's failure to maximise the good will be mitigated more than twice as much as B's.

In the Same Values Story, then, only a cost-based theory can provide more flexibility than a standard proportional theory. Such flexibility seems to be desirable in this case. It is

reasonable that, even when the amount of value at stake is constant, each successive unit of cost to the agent should count for more when weighed against that value. The Same Values Story suggests that a cost-based theory is preferable to a value-based one.

Let us turn now to the Same Costs Story. Here, the costs to the agent are the same in A's case as in B's. So any difference in wrongness must be due to differences in the amounts of value at stake. On a cost-based theory, the relationship between wrongness and value is strictly proportional. Therefore, if the costs are unaltered, then wrongness must be proportional to value. If B does z rather than w and A does y rather than x, then B does something exactly twice as wrong as what A does. In the Same Costs Story, a cost-based theory must deliver precisely the same moral judgements as a fully proportional theory.

In a value-based theory, by contrast, wrongness is not proportional to value, even when costs are unvarying. Therefore, if B does z and A does y, B will do something which is less than twice as wrong as what A does, as the 50 units of value between 50 and 100 count for more in the assessment of wrongness than those between 100 and 150. In the Same Costs Story, only a value-based theory can provide more flexibility than a standard, proportional account.

Now, if the arguments of Chapter SIX are correct, then such flexibility is indeed desirable, as our focus there was on situations such as the Same Costs Story. In such situations, a value-based theory is preferable to a cost-based one.

Investigation of the cases where the two theories diverge suggests that the best theory may be a combination of the two, one which combines the non-proportional account of value-based wrongness outlined in Chapter SIX with the non-proportional appeal to cost developed in this chapter. Such a theory would provide the optimal amount of flexibility.

However, for the remainder of this chapter I shall focus on a purely cost-based theory.

There are several reasons for this. The first is that a doubly non-proportional theory would be doubly complicated. In the interests of simplicity it is better to concentrate on only one aspect of non-proportionality at a time. Furthermore, in most situations, the two non-proportional theories will be more similar to each other than either is to any proportional Hybrid View. In practice, any increase in the amount of value at stake is likely to be accompanied by an increase in the cost to the agent, and vice versa. Therefore, on either theory, the increase in the sacrifice required from the agent will be less than proportional to the increase in the amount of value at stake, and also less than proportional to the increase in the cost involved. As we are primarily interested in finding a theory which offers smaller changes than those required by the proportional Hybrid View, focusing on one element is thus unlikely to significantly distort our analysis.¹⁵²

¹⁵²Also, I wish to suggest that the choice between the two theoretical structures is not as significant as it may seem. An agent's obligations depend upon her moral situation, which is a combination of features of the agent herself and features of the world. For instance, the amount of value at stake is primarily a feature of the world, whereas the structure of the agent's interests is primarily a feature of her. It is tempting to then regard promotion-of-good reasons as part of the agent's world, while treating cost-based reasons as intrinsic to the agent. The choice between our two versions of the Hybrid View then takes the form of a decision as to whether the non-proportional element of our account should be located with the agent, yielding a non-proportional appeal to cost, or in the world, yielding a non-proportional account of the relationship between wrongness and the reason to promote the good. As the purpose of the non-proportional element is to render our account more sensitive to the situations of particular agents, it seems more plausible to locate that element with the agent, leaving the other components factors of wrongness fully proportional. However, I think it is a mistake to associate some reasons exclusively with the agents and tie others exclusively to the world. While values may, from the point of view of a Consequentialist moral theory, be treated as being 'in the world', reasons of any kind are definitely not. Nor, on the other hand, should a Consequentialist account allow reasons to arise for the agent in isolation from the world. Indeed, it is clear that not even an agent's interests are in any strong sense independent of all the external features of her situation. Rather, reasons, as distinct from values, arise only for a particular agent in a particular situation. Reasons of all kinds are intermediate between the agent and the world. They are not facts about the world OR facts about the agent. They don't tell the agent anything about her world or about herself. Rather, an agent's reasons tell her how she should respond to her world. The reason to promote the good does not tell Affluent about the value of her world or possible future states of it. It tells her how she, in her situation, should respond to those values.

Once we see all reasons in this light, then it becomes less plausible to regard some as being located closer to the agent than others. Indeed, the whole idea of a rigid divide between two conflicting sets of reasons - reasons to promote the good versus reasons derived from the cost to the agent - comes to seem artificial. We would do better to speak of one kind of reason which is responsive both to impersonal values and to the agent's own interests. Our primary interest is in the verdict of the balance of reasons, not in the precise composition of the reasons which go to make up that balance. Of course, in most cases, different accounts of the number and types of reasons involved will produce widely different accounts of what, on the balance of reasons, a particular agent should do in a particular situation. This is why moral philosophers are right to taxonomise moral reasons. However, in many cases, the choice of where to place the non-proportional

Having decided to focus on one element, we need to decide which it will be. I choose to focus on cost primarily because non-proportional approaches to value-based wrongness were discussed extensively in the last chapter. Also, some of those who were unconvinced by that discussion may find the notion of a non-proportional appeal to cost more intuitively acceptable. It is thus worth exploring this second dimension of non-proportionality on its own.¹⁵³

Let us now summarise our scalar, non-proportional agent-centred prerogative in two basic conditions.¹⁵⁴ Before doing so, let us introduce some simplifying terminology. Let the agent's "simple reason" to do x rather than y be her promotion-of-good reason to do x rather than y, let the "weighted cost" to the agent in doing x rather than y be the cost she would have borne if she'd done x rather than y (multiplied by the weighting she's allowed to give to her own interests at that level of cost), and let her "weighted reason" to do x rather than y be her simple reason minus the weighted cost involved.

Two basic conditions of wrongness suggest themselves:

The Absolute Wrongness Condition : An act (y) is wrong if and only if there is at least one other act (call it x) available to the agent such that both (i) the agent forgoes some positive

component in our revised Hybrid View will have little or no effect whatsoever on which moral judgements our theory will produce. So, in this particular case, the technical decision between rival combinations of reasons is a sterile one which may reasonably either be left open or settled in whichever way we happen to find most convenient or congenial.

¹⁵³The focus on cost-based theories does raise one other pressing question: if the appeal to cost is able to do all the work of a non-proportional account of the relationship between wrongness and values, why bother with the latter at all? Briefly I have two justifications. The first is that the primary purpose of Chapter SIX was to introduce the notion of non-proportionality in general into our account of wrongness. The location of that element of non-proportionality was of secondary importance. Secondly, the appeal to cost itself is by no means unproblematic, while the reason to promote the good is certainly not an optional extra for Consequentialists. The order of our exposition shows us that non-proportionality can be introduced into a Consequentialism which does not include an appeal to cost.

¹⁵⁴From now on, all references to the non-proportional Hybrid View should be read as referring to this cost-based version.

amount of value in doing y rather than x; and (ii) the weighted cost the agent would have borne if she had performed x instead of y is LESS than the agent's simple reason to do x rather than y.

The Comparative Wrongness Condition : Let y and z be two acts, both available to the same agent at the same time, and both wrong according to the Absolute Wrongness Condition. Let x be an act which meets conditions (i) and (ii) for the Absolute Wrongness Condition with respect to both y and z. Then, y will be more wrong than z if and only if the agent's weighted reason to do x rather than y is GREATER than her weighted reason to do x rather than z.

7.2: Objections to the Hybrid View Reconsidered

We turn now to our second question: does our new hybrid provide an adequate response to the Demandingness Objection? We begin by asking whether or not our new theory can meet the objections we raised in Chapter FIVE against Scheffler's original Hybrid View.

7.2.1: The Wrong Facts Objection

The principal objection we must discuss is the Wrong Facts Objection. Recall the following tale from section 5.2.1 of Chapter FIVE:

Affluent's Fourth Ignorance : Affluent is sitting at her desk with her cheque book. She has heard a rumour that the effectiveness of Oxfam's operations (that is, the amount of good they can do for each dollar they receive) has been drastically altered. However, she does not know whether or not this rumour is true; or, even if it is true, what the nature of the alteration has been. To simplify, let us say that there are three possibilities:

Inefficient Oxfam : Oxfam's efficiency has declined to 10% of its previous level. That is, each dollar spent now produces only one tenth as much good as it previously would.

Normal Oxfam : Oxfam's efficiency is unchanged.

Superefficient Oxfam : the efficiency of famine relief operations has increased tenfold. Each dollar spent now produces ten times as much good as it previously would.

Our revised agent-centred prerogative enables Scheffler to deal with Affluent's Fourth Ignorance. Assume that in Normal Oxfam Affluent is required to make a 10% donation. The shift to Superefficient Oxfam increases the amount of good which a given donation will produce. Affluent will be required to bear a greater cost in Superefficient Oxfam than in Normal Oxfam. However, that cost will not have increased by a factor of ten. Similarly, in Inefficient Oxfam, Affluent's promotion-of-good reasons will be weaker, so the cost she is required to bear will be reduced. However, it will not be reduced by a factor of ten.

Let me explain why this is so. Recall that, in section 5.2.1 of Chapter FIVE, we concluded that in Normal Oxfam Affluent has given 10% of her income to Oxfam when she reaches the point at which the amount of additional utility she could gain by spending her next dollar on herself, when multiplied by the disproportionate weighting she's allowed to give to her own interests, is equal to the amount of additional value Oxfam could produce by spending that dollar.

Under the original (proportional) version of the agent-centred prerogative, Affluent must continue making donations in Superefficient Oxfam until the amount of additional utility she would gain by spending her next dollar on herself is ten times as great as it would have

been at the point at which she stopped making donations in Normal Oxfam. By contrast, under our non-proportional theory, Affluent must continue making donations in Superefficient Oxfam until the amount of additional utility she would gain by spending her next dollar on herself is p times as great as it would have been at the point at which she stopped making donations in Normal Oxfam (where $1 < p < 10$). As the amount of value at stake increases, the total cost Affluent must bear to produce that value will also increase. However, each successive unit of cost counts for more than the last. Things which are ten times as valuable to Affluent as those she was required to give up in Normal Oxfam will count for more than ten times as much. Affluent will thus not be required to give those things up. In Superefficient Oxfam, Affluent is only required to give up things which are p times as valuable to her as the things she is required to give up in Normal Oxfam, rather than those which are ten times as valuable.

Recall also that, under the original agent-centred prerogative, in Inefficient Oxfam, Affluent was only required to continue making donations until the amount of additional utility she would gain by spending her next dollar on herself was one tenth as great as it would have been at the point at which she stopped making donations in Normal Oxfam. By contrast, under our non-proportional theory, Affluent must continue making donations in Inefficient Oxfam until the amount of additional utility she would gain by spending her next dollar on herself is q times smaller than it would have been at the point at which she stopped making donations in Normal Oxfam (where $1 < q < 10$). In Inefficient Oxfam the agent-centred prerogative would allow Affluent to indulge in things which are one q -th as valuable to her as the things she would be allowed to indulge in in Normal Oxfam, rather than in those which are only one-tenth as valuable.

If we set p and q both equal to ten, then our non-proportional agent-centred prerogative will be identical to Scheffler's original version. At the other extreme, setting p and q both equal

to one would give us a theory in which the demands placed on Affluent remained exactly the same throughout our three scenarios. By setting p and q sufficiently close to one, we can ensure that Affluent's prerogatives remain similar in each of the three cases. Our initial objection was not that Scheffler made different requirements in the three situations, but rather that those differences were too great. The introduction of a suitable non-proportional agent-centred prerogative should thus be sufficient to dissolve our objection.

It seems that our new theory may be open to a variant of the Wrong Facts Objection in the following way. For any level of cost to the agent, there will be some natural number (n) which is such that the amount of good produced by saving n people from starvation is greater than the cost to the agent multiplied by the weighting the agent is allowed to give to her own interests at that level of cost. It follows that, if we set the number of people who would be saved in Superefficient Oxfam sufficiently high, we can construct a situation in which the cost required of Affluent in Superefficient Oxfam is exactly ten times that required in Normal Oxfam. We can then retell Affluent's Fourth Ignorance, replacing the original Superefficient Oxfam with this new version. Once again, Affluent will need to know whether or not Oxfam's efficiency has increased before she has any idea how much she's obliged to donate.

However, this new Wrong Facts Objection lacks the force of the original. In the first place, we should note that the Wrong Facts Objection is weaker against the Hybrid View than against either Collective Consequentialism or Murphy's view. This is because, in Affluent's Fourth Ignorance, in contrast to her first three ignorances, the amount of good Affluent is able to do with any given level of donation does vary between the different situations. I did not object in principle to the notion that such differences in the agent's ability to affect the world should impact on her obligations. I objected only to the claim that changes in obligations should be directly proportional to any such change in ability.

If we grant that variations in the agent's ability to affect the world should have some effect on the extent of her obligations, then the Wrong Facts Objection can no longer provide a general repudiation of non-proportional Hybrid Views. To see this, assume that we start out with a particular non-proportional Hybrid View which specifies the precise relationship between the cost to the agent and the weight she's allowed to give to her own interests. We are then presented with a situation like Affluent's Fourth Ignorance in which, on this theory, variations in the agent's abilities would produce an intuitively disproportionate change in her obligations. The existence of such a situation does indeed count against that particular revised Hybrid View. However, there will be countless other non-proportional hybrid theories which would not yield an unacceptable result in that particular situation. We would thus be free to replace our first theory with one of those. Of course, we know that there will be situations in which our new hybrid theory will lead us from variations in the agent's abilities to significant variations in the extent of her obligations. However, given that we've accepted some such variations, we cannot conclude in advance that our new theory produces counter-intuitive variations. It may turn out that it does, but then we can replace it with a third theory which avoids those particular variations.

The discovery of counter-intuitive variations between the agent's ability to affect the world and the extent of her obligations thus provides a means of refining a particular non-proportional hybrid theory, not an objection to such theories in general. If there is an intuitively acceptable pattern of variations, then, a fortiori, there must be some theory which countenances those, and only those, variations. The revised Wrong Facts Objection gives us no reason to believe that this ideal theory will not be some non-proportional version of the Hybrid View.

7.2.2: Kagan's Objection

In Chapter FIVE, I discussed Kagan's objection that the original Hybrid View allowed agents to kill in order to further their own personal interests. I concluded that Scheffler has yet to convince us that the Hybrid View can defeat this objection. It does not seem likely that the introduction of non-proportionality will provide the Hybrid View with any additional resources with which to meet Kagan's objection. If our non-proportional agent-centred prerogative is to form part of an acceptable overall moral theory, then we may still need to combine it with some constraint on killing for personal gain. As I noted in Chapter FIVE, an adequate assessment of such constraints lies beyond the scope of this thesis.

7.3: Conclusions

I conclude that the revised Hybrid View is superior to Scheffler's original, as it is able to dissolve the Wrong Facts Objection.

Chapter EIGHT: Blameless Wrongdoing Revisited

A second possible supplement to the account of wrongness developed in Chapter SIX would be to incorporate it into a theory which adopts Parfit's distinction between wrongness and blameworthiness (which was discussed in Chapter ONE). In this chapter, I briefly outline how such a theory might be developed, and attempt to show that it would be superior to Parfit's original account.

8.1: A Unified Account of Wrongness and Blame

In any Consequentialist theory, we are interested in comparing how the world goes with how it would have gone if an agent had done something different. So (roughly speaking), in assessing A's doing x rather than y we are comparing the actual world (in which A does x) with a possible world (W) such that:

(i) in W, A does y at t;

(ii) of all the worlds which satisfy condition (i), W is the closest to the actual world.¹⁵⁵

Our view of the nature of moral choice will affect how we characterise W. In particular, we can distinguish two contrasting views:

The Radical Choice View : W will be a world in which everything prior to t is exactly as it is in the actual world, but in which, at t, A does y. (The two worlds will then diverge after t)

¹⁵⁵This account of counterfactuals is controversial, but it will be sufficient for our present purposes. For further discussions, see Lewis 1973 and 1979, and various papers in Jackson 1991a.

The Ongoing Self View : at t, A's action is not the result of a radical, un-determined choice. Rather, it results in large part from A's character, values, motivations etc., and from certain general laws of moral psychology. A world in which everything prior to t is exactly as it is in the actual world, but in which A does y at t, would either have to be a world in which those general regularities suddenly ceased to hold at t, or one in which A's character suddenly changed at t. Such a world would not be very close to the actual world at all. The closest world in which A does y at t is thus likely to be a world which also differs from the actual world prior to t. In particular, it will be a world in which A's character is different from her character in the actual world.

In effect, we are interested in counterfactuals of the following form:

C: If A [who actually did x at t] had done y at t, then p would have been the case.

Our two competing views provide alternative analyses of C. The Radical Self View analyses C as what Parfit has dubbed the Standard Counterfactual:¹⁵⁶

Cs: If A [who actually did x at t] had done y at t, and everything prior to t was as it is in the actual world, then p would have been the case.

By contrast, the Ongoing Self View glosses C in terms of what Lewis calls a Backtracking Counterfactual:¹⁵⁷

¹⁵⁶See Parfit 1994.

¹⁵⁷Lewis 1979, p. 456.

Cb: If A [who actually did x at t] had done y at t, and if things before t were as they would need to have been for it to come about that A did y at t, then p would have been the case.

Any defence of the Ongoing Self View thus consists of a defence of the use of Backtracking Counterfactuals in such cases. Now, in the majority of situations, standard counterfactuals are clearly more relevant than backtracking counterfactuals. For most moral judgements, standard counterfactuals are to be preferred. However, I shall argue that backtracking counterfactuals are more appropriate in some moral cases.

This claim is based on a general account of subjunctive conditionals on which the notion of "closest world in which ..." is context-dependent.¹⁵⁸ Different kinds of counterfactual are appropriate in different contexts. Not all moral questions raise the same counterfactual questions. In particular, in the context of assessing an agent's character, we should appeal to backtracking counterfactuals. This is despite the fact that, as Parfit demonstrates, standard counterfactuals are more appropriate when we are assessing acts considered in isolation.¹⁵⁹ For the Consequentialist, what an agent ought to do is determined by standard counterfactuals. How she ought to be judged, however, depends on backtracking ones.

The reason for this difference is a distinction between two ways of viewing a moral situation. The first, the Act Focus, concentrates on alternatives available to the agent at time t, as things stand at t. Under this focus, an act is available to the agent so long as it is possible for the agent to perform that act. The notion of possibility here is a weak compatibilist one, on which A could have done x if and only if she would have done x if

¹⁵⁸This claim is defended in, for instance, Lewis 1979 and Tichy 1984.

¹⁵⁹Parfit 1994.

she'd wanted to.¹⁶⁰ For the Act Focus, we hold everything prior to *t* as it is in the actual world, and then ask how things would have turned out if, nonetheless, the agent had acted differently. When we're asking what *A* should have done at *t*, it would clearly be absurd to reply with an account of what *A* should have done at some time prior to *t*. For instance, if we ask how, given that she has children, Affluent should balance their needs against the requests of Oxfam, it will not do to reply simply that if Affluent hadn't had children in the first place, then she would be better able to help Oxfam now.

The second point of view is the Agent Focus. Here, our interest is in understanding and evaluating the agent's character, as it is expressed in different actions at *t*. Suppose the agent actually did *x*, and we are looking at another option (*y*). From the Agent Focus, two questions are pertinent: What sort of person would do *x* at *t*? (i.e., what sort of person is the agent?) and What sort of person would do *y* at *t*? (i.e., what sort of person would the agent have needed to be in order to have done *y* at *t*?) We are thus interested in those non-actual situations in which the agent has whatever character would best explain her doing *y* instead of *x*. In other words, we seek to tell the most plausible tale in which *A* does *y* at *t*, while departing as little as possible from the actual world. Given these interests, a backtracking counterfactual is clearly more appropriate.

We should note that, in many cases, the standard and backtracking counterfactuals will converge. For instance, consider the following two examples:

The Plausible Lapse : Mary, a usually reliable individual, is moderately security conscious. She knows that bike thefts are commonplace on campus and has always locked her bike.

¹⁶⁰We could also justify the Act Focus under a stronger contra-causal account of possibility, if we adopted the libertarian view that contra-causal freedom is widespread.

This morning, however, as she rushes to make her tutorial on free-will, Mary does not lock her bike.¹⁶¹

The Inexplicable Lapse : Mary is a sane, happy individual. While sitting in her philosophy tutorial this afternoon, she suddenly jumps out a tenth story window.

Assume that, in the actual world, Mary does lock her bike and does not jump out the window. I then tell you my two stories. In each story, nothing is varied prior to the moment of action, but Mary's act then diverges from her actual one. I would suggest that, while you would accept my first story, you would reject the second, as being one in which Mary's action cried out for an explanation.¹⁶²

These examples bring out an important feature of my use of Backtracking Counterfactuals. It is tempting to assume that Backtracking Counterfactuals require psychological determinism. This is because, if A's acts are not psychologically determined, then things already are as they need to be for it to come about that A does y at t. We could reply that psychological determinism is true, so this objection is irrelevant. However, I do not wish the claims of my thesis to be conditional on any particular resolution of the free-will debate, so I shall seek an alternative response. A better reply is that, despite appearances, Backtracking Counterfactuals do not require psychological determinism. The claim that they do is based on a weak reading of "things before t being as they would need to have been for it to come about that A did y at t". On this reading, the condition is satisfied so long as, at t, it was causally possible for A to do y. So it is met in any world where A's not doing y was not causally determined.

¹⁶¹I owe this example to Hanjo Glock.

¹⁶²Of course, if we assume some degree of contra-causal freedom, then Mary's jump may not, strictly speaking, be inexplicable. However, it clearly does require explanation in a way in which her failure to lock her bike does not.

This reading is too weak. I propose a stronger reading, on which our backtracking condition is met if, and only if, things prior to *t* are such that they render *A*'s doing *y* at *t* explicable. I would then claim that, when assessing an agent's character, we are interested in how that character "makes sense" of her actions. We are thus more interested in worlds where her actions do not cry out for explanation. It is this interest which underlies the strong reading of the backtracking condition.

Derek Parfit has suggested the following alternative formulation of the Agent Focus.¹⁶³ From this point of view, we are interested, not only in *A*'s character as this is reflected in her actions at *t*, but also in any past acts or omissions by which she could have affected her present character. When assessing our backtracking counterfactual, we limit our attention to those characters which were available to the agent at some previous time, in the sense that she could have taken steps at the time to ensure that she now had that particular character. On this account, psychological determinism would not create any special problems for the Agent Focus, as the problem of which possible characters to consider resolves into the more general question of which acts were available to the agent.

As we shall see, this alternative account may also help us to solve other problems which may affect backtracking counterfactuals. In many cases, the two accounts will yield identical results. We can thus speak ambiguously of the "best available explanation" for an agent's actions, a phrase which can be interpreted either as limiting us to explanations in terms of characters the agent could have caused herself to have, or as ranging more widely over explanations in terms of characters she might have had in more remote possible worlds. For our main purposes, there will be no urgent need to choose between these two interpretations.

¹⁶³In conversation.

If we accept that Mary's jumping requires explanation, then the appropriate analysis of the following counterfactual:

M: If Mary had jumped out the window at t , then p would have been the case.

will be a backtracking version of the following form:

Mb: If Mary had jumped out the window at t , and if things prior to t had been such as to provide an appropriate explanation for this action, then p would have been the case.¹⁶⁴

There is therefore a unity between judgements of acts and judgements of character. Both are based on assessments of counterfactuals of the form "if A had done x , then p would have been the case". The difference is that judgements of acts are based on standard counterfactuals whereas judgements of character are based on backtracking counterfactuals.

Consider an individual (A) who faces a choice at time t between three options (x , y and z). Let the backtracking value of x be the amount of value there would have been in the history of the world if A had done x at t , where the relevant counterfactual is a backtracking one. Similarly, for the backtracking values of y and of z . Let the backtracking value of x over y be the difference between the value of x and the value of y . Let x be the blameless act for A at t , where x is the blameless act if and only if, it is the case that if A had done x at t , then the history of the world would have gone at least as well as it would have if A had done anything else at t . We can then draw up the following conditions for a non-proportional account of blameworthiness:

¹⁶⁴Obviously, much turns on how we cash out the notion of "appropriate explanation". Standards of explanation may be context-dependent. However, this is not necessarily an objection to our account, as judgements of character may be similarly variable.

The Zero Blameworthiness Condition : The blameworthiness of the blameless act is zero.¹⁶⁵

The Comparative Blameworthiness Condition : The blameworthiness of z is greater than the blameworthiness of y if and only if the history of the world would go better if A did y at t than if A did z at t .

The Decreasing Blameworthiness Condition : Assume that the backtracking value of y is greater than the backtracking value of z . By the comparative condition, the blameworthiness of z is greater than the blameworthiness of y . Now, the value of x over z is equal to the backtracking value of x over y plus the backtracking value of y over z . The Decreasing Condition says that, in assessing the blameworthiness of z , each unit of backtracking value between y and z counts for more than each unit of backtracking value between x and y . In other words, even if the backtracking value of y over z is exactly equal to the backtracking value of x over y , the former will attract more blame for A than the latter. In this case, that would mean that z would be more than twice as blameworthy as y . In general, the ratio of the blameworthiness of z to the blameworthiness of y is greater than the ratio of the backtracking value of x over z to the backtracking value of x over y .

The Comparative Blameworthiness Condition ensures that an agent does something less blameworthy if she acts from a motive which produces more good. The Decreasing Blameworthiness Condition ensures that blameworthiness is not proportional to the amount

¹⁶⁵This is analogous to the account given in Chapter SIX of the zero degree of wrongness of the most beneficent act. However, because we are operating with backtracking counterfactuals rather than standard counterfactuals, the two notions will sometimes come apart. As we shall see below, the blameless act may not be the act with the best consequences.

of additional backtracking value the agent would have produced if she had done the blameless act.

8.2: Objections to the Unified Account

We now discuss some objections to this non-proportional account of blameworthiness.

8.2.1: The Past Affects the Present

If we allow things to vary before t , then won't that affect the nature of the choice faced by the agent at that time? This will be the case either if the pre- t divergence alters the actual options available to the agent, or if it significantly changes the likely consequences of those actions.¹⁶⁶

The best response to this objection is to rephrase our backtracking counterfactuals so as to avoid any undesirable effect on the choice situation at t . For instance, we can plausibly gloss "A does y at t " in such a way that the antecedent of our counterfactual will only come out true in situations in which A faces the same options at t as she actually does, and in which those options have roughly their actual consequences. If her options and/or their consequences were radically different, it wouldn't be y that A was choosing.

This rephrasing may seem ad hoc. However, our gloss of our own counterfactuals depends upon our interests. If we adopt the Agent Focus, then we are primarily interested in what sort of person would have done y at t in A's situation, not in what sort of person would have done something vaguely similar in a completely different situation. So our rephrasing brings the counterfactual in line with our interests, and is thus legitimate.

¹⁶⁶This argument is drawn from Parfit 1994.

8.2.2: The Indeterminacy of Backtracking Counterfactuals

One problem with backtracking counterfactuals is that we often have many such counterfactuals available to us, all with very different consequents, with no principled way to choose between them. In many cases, there will be no proposition (p) such that it is true that, if A had done y at t, and if things before t were as they would need to have been for it to come about that A did y at t, then p would have been the case.

Backtracking counterfactuals certainly are, in general, more indeterminate than standard ones. However, some indeterminacy is acceptable, perhaps even desirable. This is especially true when we are assessing moral character. Such assessment is a very complicated business, and a certain amount of indeterminacy is to be expected. I would suggest that, as the assessment of character is more complex than the assessment of acts, we should expect the former to involve more indeterminacy than the latter. The greater indeterminacy of backtracking counterfactuals may thus be a point in their favour, in this particular context.

Furthermore, the more factors we hold constant, the less indeterminacy there will be. For instance, holding constant the nature of the choice situation at t should make it easier to locate the closest possible world in which A does y at t. As our interest is only in what sort of person would do y at t in A's situation, we can hold a great many things constant. We can thus greatly reduce the indeterminacy of our backtracking counterfactuals.

As I noted above, in many cases (such as that of Mary's failure to lock her bike), our backtracking counterfactual will be identical to a standard counterfactual. It will therefore be no more indeterminate than a standard counterfactual. Furthermore, the cases where our

backtracking counterfactual is not identical to a standard counterfactual are precisely those situations where we might expect a more than usual degree of indeterminacy: i.e., when we are asking what would have been the case if the agent had done something so out of character that it would have cried out for explanation (such as Mary's jumping out the window). It is a strength of my account that it renders the relevant counterfactuals more indeterminate in these more controversial cases.

Even when we cannot say exactly what sort of person A would have needed to be in order for her to have done y instead of x, we will often be able to say something interesting about what her character would have been. For many purposes, this something will be all we need.¹⁶⁷

Finally, even if there are some cases in which the relevant backtracking counterfactuals are irredeemably indeterminate, this does not prove that such counterfactuals are of no use in those cases where they are less vague. Indeed, the fact that our analysis has nothing to say about an agent's character in some situations is not necessarily an objection to that analysis, even in those cases. Perhaps the situation itself is such that there is nothing worthwhile to be said, apart from what has already been said in the evaluation of the act taken in isolation. The fact that our analysis has nothing to say would then be a point in its favour.

I conclude that the indeterminacy of backtracking counterfactuals does not constitute a decisive objection to an analysis which uses them to assess moral character. However, the

¹⁶⁷We should note that standard counterfactuals are not exactly lacking in indeterminacy either. Specifying exactly what things would have been like if p had been the case is always notoriously difficult, especially if we're dealing with human behaviour. So the indeterminacy of backtracking counterfactuals cannot be an objection in principle to their use in moral contexts. Almost no everyday counterfactuals have the degree of precision required in formal logic. Fortunately, the requirements of moral philosophy are much looser.

prevalence of such indeterminacy suggests that, in practice, we may well need additional principles in order to pick out the best moral lives.¹⁶⁸

8.2.3: Previous Objections Reconsidered

In Chapter ONE, I raised two compelling objections to the claim that the notion of blameless wrongdoing could be used to defeat the Demandingness Objection. On the one hand, under Parfit's original definition, Affluent clearly is not blameless, so blameless wrongdoing cannot be appealed to in this particular case. On the other hand, examination of the Warlord's tale suggested that the idea of blameless wrongdoing is misconceived. We must now ask whether our revised account meets these objections.

Against the charge that blameless wrongdoing does not apply in Affluent's case, two things can be said. First, in our new account, blameless wrongdoing is not intended as the sole solution to the Demandingness Objection, but merely as a part of that solution. Secondly, our account of blameless wrongdoing differs from Parfit's in two ways. It is scalar and it is non-proportional. I shall argue below that these differences allow us to appeal to blameless wrongdoing in Affluent's case without having to claim that Affluent is completely blameless.

These differences between my new account and Parfit's original do not seem likely to provide me with a response to the problems raised by Warlord's Tale. Therefore, any solution which is available to me must also be available to Parfit. In Chapter ONE, I discussed a number of possible Parfitian responses, but found none which was entirely

¹⁶⁸We should note in passing that we could avoid these problems of indeterminacy altogether if we adopted Parfit's stricter account of the Agent Focus, on which we are interested only in characters which the agent could have caused herself to have. On this account, we simply look for the causally possible story in which Mary's character would have made things go best overall. The only indeterminacy which might arise would be a familiar and more general indeterminacy surrounding the values of possible outcomes.

satisfactory. At this stage, then, our overall account is incomplete. However, this failing is due to the incompleteness of the original Parfitian account, not to some new defect in our new account.

8.3: The Unified Theory Applied

The account of blameworthiness offered here is controversial and incomplete. We need to specify in more detail the criteria by which we judge one possible world to be "closer" to the actual world than another, for the purposes of moral evaluation. In general, as the wrongness of an act increases, the total utility in the history of the world in which the agent does that act will decrease. It follows that, if one act has a higher degree of wrongness than another, it will probably also have a higher degree of blameworthiness. Wrongness and blameworthiness tend to go hand in hand.

However, we are most interested in cases where the two come apart. Consider the following story:

The Revised Clare's Tale : Clare and her friend Impartial each face a choice between providing a small benefit to her own child or a larger benefit to a distant stranger. Clare, motivated by her love for her child, provides the benefit to him rather than to the stranger. Impartial, motivated by a dispassionate concern for the welfare of all individuals, benefits the stranger rather than her own child. Assume that the two friends live in a world in which, in stark contrast to the actual world, situations such as theirs are very rare, as there is virtually no poverty or starvation. Overall, the world would be a better place if parents had Clare's powerful motives rather than Impartial's dispassionate ones.

Both characters face the same options, so their situations, taken in isolation, are identical. Impartial chooses the option which produces the most value, so her action has a zero degree of wrongness. It is therefore not wrong. By contrast, Clare picks an option which is less beneficent than the most beneficent available act. So her choice is wrong. When we assess the motives which produced those choices, however, the situation is reversed. If Impartial had chosen differently in this particular instance, that would have been because she had motives similar to those of Clare. The history of the world would have gone better if she'd had those motives than it does given her actual motives, as her increased contribution to the good on other occasions would have more than outweighed her lower contribution on the present occasion. Impartial's action thus has some positive degree of blameworthiness. By contrast, if Clare had chosen differently in this particular instance, that would have been because she had motives similar to those which Impartial actually has. Yet, the history of the world would have gone worse if she'd had those motives rather than those she actually has, as her increased contribution to the good on the present occasion would have been outweighed by her lowered contribution on other occasions. Clare's action thus involves no blameworthiness.

It is therefore possible for wrongness and blameworthiness to point in opposite directions. In such cases, we can say that, when it comes to the assessment of blameworthiness, the wrongness of an act is completely mitigated by the desirable effects of the motive from which it springs. In less extreme cases, the wrongness of an act may be partially mitigated by the desirable effects of that motive. This will occur when, although that motive produces a greater contribution to the good on some occasions, these are not sufficient to outweigh its negative effects on other occasions. I shall argue below that the wrongness of not donating all one's money to charity may be at least partially mitigated in this way by the desirable effects of a motive which prompts such a refusal.

In Chapter ONE, I argued that blameless wrongdoing cannot, on its own, provide a solution to the Demandingness Objection. However, my present account differs from Parfit's original in two crucial respects. In the first place, my account of blameworthiness is combined with a stronger account of wrongness (featuring scalar, non-proportional and cost-based elements). I am therefore not asking the distinction between wrongdoing and blameworthiness to do all the work of defeating the Demandingness Objection. Secondly, we have incorporated non-proportional elements directly into the account of blameworthiness itself. I shall now explain in more detail how these elements might work.

To explore the non-proportional element in our account of blameworthiness, we should focus on of the difference between the total utility in the history of the world in which the agent does what she does and the total utility in the history of the world as it would have needed to be in order to best explain her having performed the blameless act in the present situation. Consider the following table, representing the choices faced by an agent (A) at time t:

Option	Backtracking Value
x	100
y	60
z	50

where the figures in the value column represent the total amount of value (above an arbitrary threshold of, say, ten billion units) in the history of the world in which A does that act at t, where all such counterfactuals are interpreted as backtracking counterfactuals. The blameless act is x. If A does z at t then what she does is more blameworthy than if she'd done y at t, as the backtracking value of y is greater than the backtracking value of z.

On a proportional account of blameworthiness, the difference in blameworthiness between A's doing x at t and A's doing y at t would be four times greater than the difference in blameworthiness between A's doing y at t and her doing z at t, as the differences in blameworthiness would be proportional to the differences in backtracking value.

Accordingly, y would be much closer in terms of blameworthiness to z than it was to x.

Under a non-proportional account, by contrast, the difference in backtracking value between z and y counts for more than any subsequent increment of ten units. The difference between the blameworthiness of x and that of y will thus be less than four times as great as the difference between the blameworthiness of y and that of z. Indeed, if our element of non-proportionality is strong enough, y may even be closer in terms of blameworthiness to x than it is to z. In other words, the ten units of backtracking value between z and y may count for more than all of the thirty subsequent units between y and x.

8.4: Applying the Unified Theory to Affluent's Tale

Recall Affluent's Tale. We cannot assess Affluent's act in isolation from the motives which produced it. We are particularly interested in counterfactuals of the following form:

C: If Affluent had made the optimal donation at t, then the value of the history of the world would have been V.

We wish to determine V, and thus compare the value of the history of the world where Affluent makes her actual donation with the value of the history of the closest possible world in which she makes the optimal donation. We do this by examining C. This is best done by telling stories in which Affluent makes the optimal donation at t. The simplest such story is the following:

The Standard Perfect Tale : Everything prior to t (including all the details of Affluent's motivations, character, values, commitments, etc.) is exactly as it is in the actual world. At t , Affluent takes out her cheque-book and signs all her assets over to Oxfam.

The question is whether or not Affluent's behaviour in the Standard Perfect Tale is explicable in terms of the most plausible tale we can tell about her pre- t motivations. If her actions are explicable, then our account of her blameworthiness will be identical to our account of her wrongdoing.

However, it seems to me that Affluent's behaviour in this story cries out for explanation, an explanation which is very unlikely to be forthcoming from an account of her actual pre- t motivations. Affluent's character almost certainly includes a host of dispositions and commitments which would make her sudden divestiture of assets extremely out of character. I suggest that such behaviour would be so out of character that the closest possible world in which Affluent makes the optimal donation must be one in which that behaviour is satisfactorily explained by her character in that world. No such explanation is possible in the Standard Perfect Tale. The closest possible world in which Affluent makes the optimal donation is thus not one in which everything prior to t is exactly as it is in the actual world.

We must replace our original story with something like the following:

The Backtracking Perfect Tale : Most things prior to t are exactly as they are in the actual world. However, Affluent's motives are different from her actual motives. At t , Affluent

takes out her cheque-book and signs all her assets over to Oxfam. Affluent's new motives must provide the most plausible explanation for this decision.¹⁶⁹

Let M be the motives Affluent has in the Backtracking Perfect Tale. We can then rephrase our original counterfactual as follows:

C: If Affluent had made the optimal donation at t and her motives had been M , then the total amount of utility in the history of the world would have been V .

But does this make any difference to the question of whether or not Affluent is to blame for failing to make the optimal donation? Yes, it does. To see how, we must compare the value of the history of the world in each of our two versions of the Perfect Tale. This requires a few simplifications. Assume that Affluent's life consists of only two types of situation: Donation Situations and Parental Situations. All Donation Situations are identical to the case we've been considering: Affluent must decide how much to donate to Oxfam, and the best thing for her to do is to donate as much as possible. Parental situations occur when the best thing for Affluent to do is something which she can only do if she actually feels strong affection for her children. Let us further assume that, in the actual world, Affluent performs optimally in Parental Situations (because she has strong affections for her children), but sub-optimally in Donation Situations (because she doesn't make the greatest possible donation).

In the Standard Perfect Tale, we assumed that Affluent's motives are sufficiently flexible that, although she has strong parental affections, she is able to perform optimally in Donation Situations. Those same strong affections would also enable her to perform

¹⁶⁹Where what counts as a plausible explanation is determined, *inter alia*, by our account of backtracking counterfactuals, as discussed in section 8.1 above.

optimally in Parental Situations. So, in Parental Situations, Affluent's behaviour in the Standard Perfect Tale will be identical to her behaviour in the actual world. The only difference will arise in Donation Situations, where Affluent behaves optimally in the Standard Perfect Tale and sub-optimally in the actual world. The blameworthiness of Affluent's actual behaviour will thus be simply the difference between the utility produced by making the optimal donation in a given Donation Situation and that produced by making a modest donation, multiplied by the number of Donation Situations Affluent encounters in her life. In other words, her blameworthiness would be the wrongness of her behaviour multiplied by the number of times she behaved in that way. This is why I said that, if we accept the Standard Perfect Tale, our account of Affluent's blameworthiness will parallel our account of her wrongdoing.

We turn now to the Backtracking Perfect Tale, in which Affluent's motives are different throughout. In that tale, Affluent makes the optimal donation in Donation Situations. Presumably, any set of motives which was compatible with Affluent's signing all her assets away to Oxfam would be one in which her affection for her children was not nearly as strong as it actually is. So, in the Backtracking Perfect Tale, Affluent would be unable to show genuine affection for her children in Parental Situations. She would thus be unable to behave optimally in such situations. Her degree of blameworthiness in the Backtracking Perfect Tale will thus be equivalent to the difference between the utility produced by making the optimal donation in a given Donation Situation and that produced by making a modest donation, multiplied by the number of Donation Situations Affluent encounters in her life, MINUS the difference between the utility produced by displaying genuine affection in a given Parental Situation and that produced by failing to display such affection, multiplied by the number of Parental Situations Affluent encounters in her life. Affluent's blameworthiness is thus less than the wrongness of her behaviour multiplied by the number of times she behaved in that way. Under the Backtracking Perfect Tale,

Affluent is thus less blameworthy than she was under the Standard Perfect Tale. This is because, in the latter tale, wrongness and blameworthiness come apart.

When we are judging blameworthiness, Affluent's wrongdoing is off-set by the desirability of the motive from which that wrongdoing springs. This desirability consists in the fact that, if Affluent had lacked the motive in question, she would have behaved less well at other times than she actually does.

The more Parental Situations Affluent encounters, and the greater the benefits of a given display of parental affection, the greater the degree to which her wrongdoing will be mitigated and, hence, the less blameworthy she will be.

Of course, as I argued in Chapter ONE, it is very unlikely that the benefits of displaying affection in Parental Situations will completely outweigh the costs of failing to make optimal donations in Donation Situations. The world would still have been a better place if Affluent had had the motives she would need to have had to donate all her assets to Oxfam in the Backtracking Perfect Tale. She is thus not totally blameless. However, she is more blameless than she would have been under traditional Consequentialism.

8.5 Conclusions

I conclude that our non-proportional account of blameless wrongdoing is superior to Parfit's original, and is thus able to play a more significant role in our overall response to the Demandingness Objection. However, like the former, the latter cannot provide an adequate response to the Demandingness Objection on its own.

CONCLUSION

We have been seeking a Consequentialist moral theory which does not make unreasonable demands. Many of our conclusions have been negative. In Part ONE, we examined a wide range of contemporary forms of Consequentialism, focusing on the theories of Parfit, Brandt, Hooker, Murphy, Slote and Scheffler. We found that none of these provided an adequate response to the Demandingness Objection.

In Part TWO, I presented a series of new responses to the Demandingness Objection, built around the notion of non-proportionality. In Chapter SIX, I introduced that device, and argued that it is superior to proportional forms of Consequentialism in a number of respects. I also demonstrated the desirability of introducing a scalar element into our theory. In Chapter SEVEN, I showed that a non-proportional version of Scheffler's Hybrid View would be superior to the original, as it would enable the latter to avoid the Wrong Facts Objection, which was raised in Chapter FIVE. In Chapter EIGHT, I showed that the introduction of non-proportionality would also enable the notion of blameless wrongdoing to play a more significant role in our response to the Demandingness Objection.

The discussion in Part TWO has been too sketchy to leave us with a detailed moral theory. Also, the mere introduction of non-proportionality will not enable us to avoid all our objections to previous theories. The standard non-proportional Consequentialism of Chapter SIX is still open to the objection (leveled in Part ONE against both Collective Consequentialism and Satisficing Consequentialism) that it cannot account for the moral significance of the costs faced by the agent. The non-proportional Hybrid View of Chapter SEVEN, while meeting the Wrong Facts Objection, has no additional resources with which to combat Kagan's objection that Hybrid Views must allow agents to harm others in pursuit

of their own personal interests, even at the expense of the common good. Finally, even a non-proportional version of the notion of blameless wrongdoing cannot, by itself, provide an adequate response to the Demandingness Objection.

However, we can conclude both that the notion of non-proportionality is worthy of further exploration, and that the best possible Consequentialist moral theory is very likely to incorporate some non-proportional elements. We can also conclude that it would be premature to assume that Consequentialists will never be able to put together adequate response to the Demandingness Objection.

BIBLIOGRAPHY

Adams, R. 1976, "Motive Utilitarianism", *Journal of Philosophy* 73, pp. 467-81.

Aiken, W. 1990, "Famine and Distribution", *Journal of Philosophy* 87, pp. 642-643.

Alexander, L. 1987, "Scheffler on the Independence of Agent-Centred Prerogatives from Agent-Centred Restrictions", *Journal of Philosophy* 84, pp. 277-83.

Altham, J. E. J. 1988, "Understanding the Logic of Obligation", *Proceedings of the Aristotelian Society*, supplementary volume 62, pp. 271-283.

Baier, K. 1958, *The Moral Point of View*, Cornell: Cornell University Press.

Bennett, J. 1966, "Whatever the Consequences", *Analysis* 26, pp. 83-102.

Bennett, J. 1989, "Two Departures from Consequentialism", *Ethics* 100, pp. 54-66.

Bentham, J., *Introduction to the Principles of Morals and Legislation*, Oxford: Blackwell, 1948.

Brandt, R. B. 1984, "Utilitarianism and Moral Rights", *Canadian Journal of Philosophy* 14, pp. 1-20.

Brandt, R. B. 1992a, *Morality, Utilitarianism, and Rights*, Cambridge: Cambridge University Press.

Brandt, R. B. 1992b, "Some merits of one form of rule-utilitarianism", in his *Morality, Utilitarianism, and Rights*, Cambridge: Cambridge University Press, pp. 111-136.

Brandt, R. B. 1992c, "Fairness to Indirect Optimific Theories in Ethics", in his *Morality, Utilitarianism, and Rights*, Cambridge: Cambridge University Press, pp. 137-157.

Brandt, R. B. 1992d, "Two concepts of utility", in his *Morality, Utilitarianism, and Rights*, Cambridge: Cambridge University Press, pp. 158-175.

Brink, D. O. 1986, "Utilitarian Morality and the Personal Point of View", *Journal of Philosophy* 83, pp. 417-438.

Brink, D. O. 1988, "Sidgwick's Dualism of Practical Reason", *Australasian Journal of Philosophy* 66, pp. 291-307.

Broad, C. D. 1916, "On the Function of False Hypotheses in Ethics", *International Journal of Ethics* 26, pp. 377-97. (reprinted in *Broad's Critical Essays in Moral Philosophy*, D. R. Chevey (ed.), London: Allen and Unwin, pp. 43-62).

Broad, C. D. 1930, *Five Types of Ethical Theory*, London: Kegan Paul.

Broad, C. D. 1971a, *Broad's Critical Essays in Moral Philosophy*, D. R. Chevey (ed.), London: Allen and Unwin.

Broad, C. D. 1971b, "The Doctrine of Consequences in Ethics", in *Broad's Critical Essays in Moral Philosophy*, D. R. Chevey (ed.), London: Allen and Unwin, pp. 17-43.

Broad, C. D. 1971c, "Conscience and Conscientious Action", in *Broad's Critical Essays in Moral Philosophy*, D. R. Chevey (ed.), London: Allen and Unwin, pp. 136-155.

Broad, C. D. 1971d, Critical Notice of Julian Huxley, *Evolutionary Ethics*, in *Broad's Critical Essays in Moral Philosophy*, D. R. Chevey (ed.), London: Allen and Unwin, pp. 156-187.

Broad, C. D. 1971e, "Some Reflections on Moral Sense Theories in Ethics", in *Broad's Critical Essays in Moral Philosophy*, D. R. Chevey (ed.), London: Allen and Unwin, pp. 188-222.

Broad, C. D. 1971f, "Self and Others", in *Broad's Critical Essays in Moral Philosophy*, D. R. Chevey (ed.), London: Allen and Unwin, pp. 262-282.

Carens, J. 1981, *Equality, Moral Incentives, and the Market: An Essay in Utopian Politico-Economic Theory*, Chicago: University of Chicago Press.

Carson, R. 1991, "A note on Hooker's Rule-Consequentialism", *Mind* 100, pp. 117-21.

Chisholm, R. 1963, "Contrary to Duty Imperatives and Deontic Logic", *Analysis* 24, pp. 33-36.

Conee, E. 1983, Review of Regan, D., *Utilitarianism and Cooperation*, *Journal of Philosophy* 80, pp. 415-424.

Crisp, R. 1990, "Sidgwick and Self-Interest", *Utilitas* 2, pp. 267-280.

Crisp, R. 1992, 'Utilitarianism and the Life of Virtue', *The Philosophical Quarterly* 42, pp. 139-160.

Crisp, R., "Promoting the Good and the Separateness of Persons", unpublished manuscript.

Cullity, G. 1994, "International Aid and the scope of kindness", *Ethics*, forthcoming.

Dancy, J. 1994, "Parfit and Indirectly Self-defeating Theories", in Dancy, J. (ed.), *Reading Parfit*, Oxford: Basil Blackwell, forthcoming.

Darwall, S. 1984, Review of Scheffler, S., *The Rejection of Consequentialism*, *Journal of Philosophy* 81, pp. 220-226.

Darwall, S. 1986, "Rational Agent, Rational Act", *Philosophical Topics* 14, pp. 33-57.

DeCew, J. W. 1981, "Conditional Obligation and Counterfactuals", *Journal of Philosophical Logic* 10, pp. 55-72.

Den Uyl, D. 1992, "Teleology and Agent-centredness", *Monist* 75, pp. 14-33.

- Donagan, A. 1992, "Sidgwick and Whewellian Intuitionism", in Schultz, B. (ed.), *Essays on Henry Sidgwick*, Cambridge: Cambridge University Press, pp. 123-142.
- Dreze, J., and Sen, A. K. 1989, *Hunger and Public Action*, Oxford: Clarendon Press.
- Feinberg, J. 1967, "The Forms and Limits of Utilitarianism", *Philosophical Review* 76, pp. 368-381.
- Feinberg, J. 1980, "The Nature and Value of Rights", reprinted in *Rights, Justice and the Bounds of Liberty*, Princeton: Princeton University Press.
- Fishkin, J. 1982, *The Limits of Obligation*, New Haven: Yale University Press.
- Foot, P. 1983, "Moral Realism and Moral Dilemmas", *Journal of Philosophy* 80, pp. 379-398.
- Foot, P. 1985, "Utilitarianism and the Virtues", *Mind* 94, pp. 196-209.
- Frey, R. G. 1977, "Act-Utilitarianism: Sidgwick or Bentham or Smart", *Mind* 86, pp. 95-100.
- Gibbard, A. 1965, "Rule Utilitarianism: A Merely Illusory Alternative?", *Australasian Journal of Philosophy* 43, pp. 211-220.
- Gibbard, A. 1982, "Utilitarianism and Human Rights", *Social Philosophy and Policy* 1, pp. 92-102.

Glover, J. 1975, "It makes no difference whether or not I do it", *Proceedings of the Aristotelian Society*, supplementary volume 49, pp. 171-190.

Gomberg, P. 1986, "Self and Others in Bentham and Sidgwick", *History of Philosophy Quarterly* 3, pp. 437-448.

Greenspan, P. S. 1975, "Conditional Oughts and Hypothetical Imperatives", *Journal of Philosophy* 72, pp. 259-276.

Griffin, J. 1986, *Well-Being*, Oxford: Clarendon Press.

Griffin, J. 1990, Review of Kagan, S., *The Limits of Morality*, *Mind* 99, pp. 128-131.

Hardin, G. 1977, "Lifeboat Ethics: The case against helping the poor", in *Morality and World Hunger*, W. Aiken and H. La Follette (eds.), Englewood Cliffs, New Jersey: Prentice-Hall.

Hardin, R. 1982, *Collective Action*, Baltimore: John Hopkins University Press.

Hardin, R. 1988, *Morality Within the Limits of Reason*, Chicago: University of Chicago Press.

Hardin, R. 1992, "Common-sense at the foundations", in Schultz, B. (ed.), *Essays on Henry Sidgwick*, Cambridge: Cambridge University Press, pp. 143-160.

Hare, R. M. 1976, "Ethical Theory and Utilitarianism", in *Contemporary British Philosophy - Personal Statements*, Fourth Series, ed. H. D. Lewis, London: Allen and Unwin.

Hare, R. M. 1981, *Moral Thinking*, Oxford: Clarendon Press.

Harris, G. W. 1989, "A Paradoxical Departure from Consequentialism", *Journal of Philosophy* 86, pp. 90-102.

Harris, J. 1974, "Williams on Negative Responsibility and Integrity", *Philosophical Quarterly* 24, pp. 265-73.

Harris, J. 1974, "The Survival Lottery", *Philosophy* 50, pp. 81-87.

Hooker, B. 1990, "Rule Consequentialism", *Mind* 99, pp. 67-77.

Hooker, B. 1991a, "Rule-Consequentialism and Demandingness: A Reply to Carson", *Mind* 100, pp. 270-76.

Hooker, B. 1991b, "Brink, Kagan and Self-Sacrifice", *Utilitas* 3, pp. 263-73.

Hooker, B. 1993a, Review of Scheffler, S., *Human Morality*, *Mind* 102, pp. 390-394..

Hooker, B. 1993b, Review of Nagel, T., *Equality and Partiality*, *Philosophical Quarterly* 43, pp. 366-372.

- Hooker, B. 1993c (ed.), *Rationality, rules and utility: new essays on the moral philosophy of Richard B. Brandt*, Boulder, Colorado: Imprint.
- Hooker, B. 1994a, "Rule-Consequentialism, Incoherence, Fairness", *Proceedings of the Aristotelian Society* 95, pp. 19-35.
- Hooker, B. 1994b, "Compromising with Convention", *American Philosophical Quarterly* 31, pp. 311-317.
- Howard-Snyder, F. 1992, "Elbow-room for Consequentialists", *Analysis* 52, pp. 249-253.
- Hurka, T. 1990, 'Two Kinds of Satisficing,' *Philosophical Studies* 59, pp. 107-111.
- Hurka, T. 1993, *Perfectionism*, Oxford: Oxford University Press.
- Jackson, F. 1986, "A Probabilistic Approach to Moral Responsibility", *Proceedings of the Seventh International Congress of Logic, Methodology and Philosophy of Science*, R. Bacan Marcus et al. (eds.), Amsterdam: Elsevier Science Publishers, pp. 351-366.
- Jackson, F. 1987, "Group Morality", in P. Pettit et al. (eds.), *Metaphysics and Morality*, Oxford: Blackwell, pp. 91-110.
- Jackson, F. 1988, "Understanding the Logic of Obligation", *Proceedings of the Aristotelian Society*, supplementary volume 62, pp. 255-270.
- Jackson, F. 1991a (ed.), *Conditionals*, Oxford: Oxford University Press.

Jackson, F. 1991b, "Decision-theoretic Consequentialism and the Nearest and Dearest Objection", *Ethics* 101, pp. 461-482.

Jackson, F. 1994, "Which Effects?", in Dancy, J. (ed.), *Reading Parfit.*, Oxford: Basil Blackwell, forthcoming.

Jackson, F., & Pargetter, R. 1986, "Oughts, Options and Actualism", *Philosophical Review* 95, 1986, pp. 233-255.

Kagan, S. 1984, "Does Consequentialism demand too much?", *Philosophy and Public Affairs* 13, pp. 239-54.

Kagan, S. 1987, "Donagan on the sins of Consequentialism", *Canadian Journal of Philosophy* 17, pp. 643-54.

Kagan, S. 1989, *The Limits of Morality*, Oxford: Oxford University Press.

Kagan, S. 1994, "Defending Options", *Ethics* 104, pp. 33-351.

Kavka, G. 1975, "Extensional Equivalence and Utilitarian Generalisation", *Theoria* 41, pp. 125-147.

Kitcher, P. 1985, *Vaulting Ambition: Sociobiology and the Quest for Human Nature*, Cambridge: MIT Press.

Lewis, D. 1968, "Counterpart Theory and Quantified Modal Logic", *Journal of Philosophy* 65, pp. 113-126.

Lewis, D. 1969, *Convention*, Cambridge, Mass.: Harvard University Press.

Lewis, D. 1973, *Counterfactuals*, Oxford: Basic Blackwell.

Lewis, D. 1979, "Counterfactual Dependence and Time's Arrow", *Nous* 13, pp. 455-476.

Lewis, D. 1981, "Causal Decision Theory", *Australasian Journal of Philosophy* 59, pp. 5-30.

Lewis, D. 1986, "A Subjectivist's Guide to Objective Chance", in *Philosophical Papers: Volume Two*, Oxford: Oxford University Press, pp. 83-132.

Locke, D 1987 "The Parfit Population Problem", *Philosophy* 62, pp. 131-157.

Lucas, G. R. 1976, "The Political and Economic Dimensions of Hunger", in *Lifeboat Ethics: The Moral Dilemmas of Hunger*, Lucas and Ogletree (eds.), New York: Harper and Row, pp. 1-28.

Lucas, G. R. 1990, "African Famine: New Economic and Ethical Perspectives", *Journal of Philosophy* 87, pp. 629-641.

Lyons, D. 1965, *Forms and Limits of Utilitarianism*, Oxford: Clarendon Press.

Margolis, J. 1965, "Rule Utilitarianism", *Australasian Journal of Philosophy* 43, pp. 220-225.

Mellor, H. 1983, "Objective Decision Making", *Social Theory and Practice* 9, pp. 289-309.

Mendola, J. 1986, "Parfit on directly collectively self-defeating moral theories", *Philosophical Studies* 50, pp. 153-165.

Mill, J. S., *Utilitarianism, Liberty and Representative Government*, London: Dent, 1948.

Miller, F., and Sartorius, R. 1979, "Population Policy and Public Goods", *Philosophy and Public Affairs* 8, pp. 148-174.

Moore, A. 1993, "The Utilitarian Ethics of R. B. Brandt", *Utilitas* 5, pp. 301-310.

Moore, A. 1994, "Consequentialism and Partiality", unpublished manuscript.

Moser, P. K. 1991, "Consequentialism and Self-Defeat", *The Philosophical Quarterly* 43, pp. 82-85.

Mott, P. 1973, "On Chisholm's Paradox", *Journal of Philosophical Logic* 2, pp. 197-211.

Mulgan, T. 1990, "Consequentialism", unpublished honours dissertation, University of Otago.

- Mulgan, T. 1993a, "Slote's Satisficing Consequentialism", *Ratio* 6, pp. 121-134.
- Mulgan, T. 1993b, "The Unhappy Conclusion and the Life of Virtue", *The Philosophical Quarterly* 43, pp. 357-359.
- Mulgan, T. 1994, "Rule Consequentialism and famine", *Analysis* 54, pp. 187-192.
- Murphy, L. 1993, "The Demands of Beneficence", *Philosophy and Public Affairs* 22, pp. 267-292.
- Nagel, T. 1970, *The Possibility of Altruism*, Oxford University Press.
- Nagel, T. 1979a, "Moral Luck", in *Mortal Questions*, Cambridge: Cambridge University Press, pp. 24-38.
- Nagel, T. 1979b, "War and Massacre", in *Mortal Questions*, Cambridge: Cambridge University Press, pp. 53-74.
- Nagel, T. 1986, *The View from Nowhere*, Oxford: Oxford University Press.
- Nagel, T. 1988, "Autonomy and Deontology", in Scheffler, S., (ed.), *Consequentialism and its Critics*, Oxford: Clarendon Press.
- Nagel, T. 1991, *Equality and Partiality*, Oxford: Oxford University Press.
- Naruse, J. F. 1976, "Utilitarianism, Group Action and Coordination", *Nous* 10, pp. 173-194.

O'Neill, O. 1986, *Faces of Hunger: An Essay in Poverty, Justice and Development*, London: Unwin Hyman.

Parfit, D. 1982, "Future Generations: Further Problems", *Philosophy and Public Affairs* 11, pp. 113-172.

Parfit, D. 1984, *Reasons and Persons*, Oxford: Clarendon Press.

Parfit, D. 1986a, "Comments", *Ethics* 96, pp. 832-872.

Parfit, D. 1986b, "Overpopulation and the Quality of Life", in P. Singer (ed.), *Applied Ethics*, Oxford: Oxford University Press, pp. 145-164.

Parfit, D. 1987, "A Reply to Sterba", *Philosophy and Public Affairs* 16, pp. 193-194.

Parfit, D. 1994, "Reply to Dancy", in Dancy, J. (ed.), *Reading Parfit*, Oxford: Basil Blackwell, forthcoming.

Pettit, P. 1984, "Satisficing Consequentialism", *Proceedings of the Aristotelian Society*, supplementary volume 58, pp. 165-176.

Pettit, P. 1986, "Slote on Consequentialism", *Philosophical Quarterly* 36, pp. 399-412.

Pettit, P. 1989, "Consequentialism", in P. Singer (ed.), *A Companion to Ethics.*, Oxford: Basil Blackwell, pp. 230-240.

Pettit, P. 1989, "Decision Theory and Folk Psychology", in M. Bacharach and S. Hurley (eds.), *Essays in the Foundations of Decision Theory*, Oxford: Basil Blackwell.

Pettit, P., and Brennan, G. 1986, "Restrictive Consequentialism", *Australasian Journal of Philosophy* 64, pp. 438-455.

Pettit, P., and Smith, M. 1993, "Brandt on Self-Control", in Hooker, B. (ed.), *Rationality, rules and utility: new essays on the moral philosophy of Richard B. Brandt*, Boulder, Colorado: Imprint.

Pettit, P., and Sugden, R. 1989, "The Backward Induction Paradox", *Journal of Philosophy* 86, pp. 169-182.

Prior, A. N. 1968, "The Consequences of Actions", in his *Papers on Time and Tense*, Oxford: Oxford University Press, pp. 51-58.

Rabinowicz, W. 1989, "Act-Utilitarian Prisoner's Dilemmas", *Theoria* 55, pp. 1-44.

Railton, P. 1986, "How Thinking about Character and Utilitarianism might lead to rethinking the character of Utilitarianism", *Midwest Studies in Philosophy* 13, pp. 398-416.

Railton, P. 1984, "Alienation, Consequentialism and Morality", *Philosophy and Public Affairs* 13, pp. 134-171.

Raz, J. 1986, *The Morality of Freedom*, Oxford: Clarendon Press.

Regan, D. 1980, *Utilitarianism and Cooperation*, Oxford: Clarendon Press.

Regan, D. 1983, "Reply to Sen", *Philosophy and Public Affairs* 12, pp. 113-132.

Ross, W. D. 1930, *The Right and the Good*, Oxford: Clarendon Press.

Ross, W. D. 1939, *Foundations of Ethics*, Oxford: Clarendon Press.

Sartre, J.-P. 1956, *Being and Nothingness: An essay on phenomenological ontology*, New York: Philosophical Library.

Scheffler, S. 1982, *The Rejection of Consequentialism*, Oxford: Clarendon Press.

Scheffler, S. 1985, "Agent-centres, restrictions, rationality and the virtues", *Mind* 94, pp. 409-419.

Scheffler, S. 1986, "Morality's demands and their limits", *Journal of Philosophy* 83, pp. 531-537.

Scheffler, S. 1988 (ed.), *Consequentialism and its Critics*, Oxford: Clarendon Press.

Scheffler, S. 1989, "Deontology and the agent: reply to Bennett", *Ethics* 100, pp. 67-76.

Scheffler, S. 1992a, "Prerogatives without Restrictions", *Philosophical Perspectives* 6, pp. 377-397.

Scheffler, S. 1992b, "Naturalism, Psychoanalysis and Moral Motivation", in J. Hopkins (ed.), *Psychoanalysis, Mind and Art*, Oxford: Basil Blackwell.

Scheffler, S. 1992c, *Human Morality*, Oxford: Oxford University Press.

Schneewind, J. B. 1977, *Sidgwick's Ethics and Victorian Moral Philosophy*, Oxford: Clarendon Press.

Schneewind, J. B. 1992, "Sidgwick and the Cambridge Moralists", in Schultz, B. (ed.), *Essays on Henry Sidgwick*, Cambridge: Cambridge University Press, pp. 93-121.

Schultz, B. (ed.) 1992, *Essays on Henry Sidgwick*, Cambridge: Cambridge University Press.

Seanor, D., and Foton, N. 1988, *Hare and Critics: Essays on Moral Thinking*, Oxford: Clarendon Press.

Sen, A. K. 1981, *Poverty and Famine: An Essay on Entitlement and Deprivation*, Oxford: Oxford University Press.

Sen, A. K. 1982, "Rights and Agency", *Philosophy and Public Affairs* 11, pp. 3-39.

Sen, A. K. 1983, "Evaluator Relativity and Consequential Evaluation", *Philosophy and Public Affairs* 12, pp. 113-132.

Sen, A. K. 1988, "Property and Hunger", *Economics and Philosophy* 4, pp. 57-68.

Sher, G. 1983, "Antecedentialism", *Ethics* 94, pp. 6-17.

Sidgwick, H. 1876, "The Theory of Evolution in its Application to Practice", *Mind* 1 (old series), pp. 52-67.

Sidgwick, H. 1880, "Mr Spencer's Ethical System", *Mind* 5 (old series), pp. 216-226.

Sidgwick, H. 1886a, *Outlines of the History of Ethics*, London: Macmillan.

Sidgwick, H. 1886b, *Lectures of the Ethics of T. H. Green, H. Spencer, and J. Martineau*, London: Macmillan.

Sidgwick, H. 1892, Critical Notice of Spencer, H., *Justice: Being Part IV of the Principles of Ethics*, *Mind* 1, pp. 107-118.

Sidgwick, H. 1907, *The Methods of Ethics*, Indianapolis: Hackett Publishing Company, 7th edition.

Simon, H. 1976, *Administrative Behaviour*, New York: Free Press, third edition.

Simpton, P. 1991, "Justice, Scheffler and Consequentialism", *American Catholic Philosophical Quarterly* 65, pp. 203-211.

Singer, M. G. 1992, "Sidgwick and Nineteenth Century British Ethical Thought", in Schultz, B. (ed.), *Essays on Henry Sidgwick*, Cambridge: Cambridge University Press, pp. 65-91.

Singer, P. 1972, "Famine, Affluence and Morality", *Philosophy and Public Affairs* 1, pp. 229-243.

Singer, P. 1977, "Utility and the Survival Lottery", *Philosophy* 52, pp. 218-22.

Singer, P. 1979, *Practical Ethics*, Cambridge: Cambridge University Press.

Skorupski, J. 1989, *John Stuart Mill*, London: Routledge.

Slote, M. 1985, *Commonsense Morality and Consequentialism*, London: Routledge and Kegan Paul.

Slote, M. 1991, *Beyond Optimizing: a study of rational choice*, London: Harvard University Press.

Smart, J. J. C. 1967, "Extreme and Restricted Utilitarianism", in *Theories of Ethics*, P. Foot (ed.), Oxford: Oxford University Press.

Smart, J. J. C., and Williams, B. 1973, *Utilitarianism: For and Against*, Cambridge: Cambridge University Press.

Smart, J. J. C. 1977, "Benevolence is an Over-riding Attitude", *Australasian Journal of Philosophy* 55, pp. 127-135.

Sosa, D. 1993, "Consequences of Consequentialism", *Mind* 102, pp. 101-122.

Spencer, H. 1851, *Social Statics*, London: Williams and Norgate.

Spencer, H. 1862, *First Principles*, London: Williams and Norgate.

Spencer, H. 1879, *The Data of Ethics*, London: Williams and Norgate.

Spencer, H. 1881, "Replies to Criticisms on The Data of Ethics", *Mind* 6 (old series), pp 82-98.

Spencer, H. 1862, *First Principles*, London: Williams and Norgate.

Stewart, R. M. 1993, "Agent-Relativity, Reason and Value", *Monist* 76, pp. 66-80.

Stocker, M. 1973a, "Act and Agent Evaluations", *Review of Metaphysics* 27, pp. 42-61.

Stocker, M. 1973b, "Rightness and Goodness: Is there a difference?", *American Philosophical Quarterly* 10, pp. 87-98.

Stocker, M. 1973c, "The Schizophrenia of Modern Ethical Theories", *Journal of Philosophy* 73, pp. 453-466.

Swanton, C. 1993, "Satisficing and Virtue", *Journal of Philosophy* 90, pp. 33-48.

Tichy, P. 1984, "Subjunctive Conditionals: Two Parameters "Vs" Three", *Philosophical Studies* 45, pp. 147-180.

von Wright, G. H. 1969, *An Essay on Deontic Logic and the General Theory of Action*, Amsterdam: North-Holland.

Waldron, J. 1994, "Kagan on Requirements: Mill on Sanctions", *Ethics* 104, pp. 310-324.

Williams, B. 1976, "Utilitarianism and Moral Self-Indulgence", in *Contemporary British Philosophy*, H. D. Lewis (ed.), London: Allen & Unwin.

Williams, B. 1981, *Moral Luck*, Cambridge: Cambridge University Press.

Wolf, S. 1982, "Moral Saints", *Journal of Philosophy* 79, pp. 419-439.

