

OXFORD



Shelly Kagan
The Limits of Morality

Oxford Ethics Series

Series Editor: Derek Parfit, All Souls College, Oxford

The Limits of Morality

This page intentionally left blank

The Limits of Morality

Shelly Kagan

CLARENDON PRESS · OXFORD



Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai

Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata

Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi

São Paulo Shanghai Taipei Tokyo Toronto

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States by

Oxford University Press Inc., New York

© Shelly Kagan 1989

The moral rights of the authors have been asserted

Database right Oxford University Press (maker)

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose this same condition on any acquirer

ISBN 0-19-823916-5

For Peter Harvey

Everything has been said before, but since nobody listens we have to keep going back and beginning all over again.

André Gide

Contents

Preface	xi
Acknowledgments	xiii
1. Against Ordinary Morality	1
Dramatis Personae	1
Moral Methodology	11
Promoting the Good	15
Resisting Options to Do Harm	19
Constraints	24
The Two-level Conception of Morality	32
Contracts	39
Promissory Notes	46
2. The Structure of Ordinary Morality	47
The Existence of the Pro Tanto Reason	47
The Concept of the Good	56
Moral Requirement	64
Ordinary Morality	71
The Neutrality of the Framework	80
3. Doing Harm	83
Drawing the Line	83
Paralysis	87
Interfering	92
Difficult Cases	101
Aid	106
The Relevance of the Distinction	111
Three Mysteries	121
4. Intending Harm	128
Harm Intended as Means	128
Self-defense	132
Shields	138
Permission	144
Due Proportion	151
Relevance Again	165
More Mysteries	177

5. Without Constraints	183
Abandoning the Constraint	183
The Neo-moderate	186
Avoiding the Constraint	194
Avoiding the Distinction	201
6. Avoiding the Appeal	204
Granting Constraints	204
The Self-constraint Argument	206
Rights	216
The Elements of Rights	219
Positive Rights	226
7. The Appeal to Cost	231
The Advantages of the Appeal	231
The Sacrifice of Interests	233
Other Options	241
The Neighborhood of the Account	251
The Personal Point of View	258
Reflecting the Nature of Persons	262
8. The Negative Argument	271
Two Attitudes about the Nature of Persons	271
Prudence	279
Vivid Belief	283
Others	291
Hypothetical Judges	300
Abandoning Universality	307
The Core of the Bias	314
Overcoming the Bias	321
9. The Positive Argument	331
Endorsing the Subjective Standpoint	331
The Patient's Point of View	338
Grounding Constraints	343
The Moral Point of View	350
Merits of the Subjective	357
Avoiding Moral Decisiveness	369
Noninsistent Reasons	378
The Defense of Options	381

10. Extraordinary Morality	386
The Value of Being Moral	386
Political and Social Structures	393
Freedom	399
The Demands of Morality	402
References	405
Index	407

This page intentionally left blank

Preface

Morality can be thought of as involving two different kinds of limits. Or rather, morality as it is commonly conceived—ordinary, commonsense morality—can be usefully thought of in this way. On the one hand (or so it is thought) morality imposes certain limits on our actions, ruling out various kinds of acts—e.g., harming the innocent—even if greater good might be brought about by an act of the kind in question. Limits of this first kind are imposed *by* morality. But there are also what we might think of as limits imposed *on* morality—for it is typically believed that there are limits to what morality can demand of us. Thus, it is generally held that although morality does sometimes require us to make sacrifices for the sake of others, we are not morally required to make our greatest possible contributions to the overall good. There is a limit to moral requirement.

This book is concerned with the limits of morality in both of these senses. Or perhaps I should say that it is concerned with the *purported* limits of morality—for one of the central aims of my argument will be to call into question the widespread belief that morality genuinely has limits of either of these kinds. Of course, it is undeniable that both sorts of limits have a significant place in our ordinary, commonsense moral views; they constitute basic features of (what I will call) *ordinary morality*. But I will argue that these two fundamental features of ordinary morality cannot be adequately defended.

In criticizing ordinary morality on this score I will be, in effect, indirectly defending consequentialism, the view which holds (roughly) that agents are morally required to perform the act that will lead to the best results overall. For consequentialism incorporates neither of the two kinds of limits I have described, and critics of this view have often objected to it on precisely these grounds. If consequentialism is correct there are no limits of the first kind for, in principle, any sort of act at all might be permissible in the right circumstances, provided only that it leads to the best consequences overall. And there are no limits of the second kind, for there is simply no limit to the sacrifices that an agent might be required to

make in the pursuit of the greater good. Advocates of ordinary morality have, accordingly, complained that consequentialism permits too much (permitting acts that should not be permitted) and they have complained that it demands too much (requiring sacrifices that are not in fact morally required). These are, I believe, the two most intuitively forceful objections to consequentialism.

Most discussion of consequentialism has focused on the first objection—i.e., that it permits too much. This is somewhat surprising, for in practical terms consequentialism may not differ in this area all that much from ordinary morality. Killing the innocent, e.g., will generally not have the best results overall, and so consequentialism and ordinary morality will typically be alike in forbidding it. And in many complex cases it is often unclear what act will lead to the best results, and so unclear whether consequentialism actually diverges in that case from ordinary morality. In contrast, the second objection—which turns on whether there is a limit to the sacrifices that morality can demand of an agent—indicates an area in which consequentialism and ordinary morality diverge sharply and undeniably. For consequentialism is *far* more demanding than ordinary morality in terms of the sacrifices that must be made for the greater good. In practical terms it matters enormously which view is correct. Yet this issue has received almost no careful discussion.

My driving concern in undertaking this work is with this second issue—the question of whether there are limits to the sacrifices that morality can demand of us. But as we shall see, the discussion of the second quickly necessitates discussion of the first—the question of whether certain kinds of acts are morally off-limits, even when they are necessary for promoting the overall good. As I have already indicated, I will be arguing that neither sort of limit can be successfully defended.

This work, then, constitutes a sustained attack on two of the most fundamental features of ordinary morality. This provides, in turn, indirect support for consequentialism. However, except for a few relevant comments at the end, the book does not in itself offer anything like a *positive* defense of consequentialism (something I hope to provide in a later work). My aim here is only to gain consequentialism a more sympathetic hearing by driving home the inadequacies of the common view—inadequacies that we generally overlook.

Acknowledgments

This book began as a conversation in the hallway of Russell House, at Wesleyan University, when Peter Harvey suggested to me that the really troubling question about utilitarianism was whether or not it was too demanding. That was about a dozen years ago, and since then the question that Peter first raised has never stopped bothering me for very long.

Eventually, at Princeton, I wrote a dissertation about it. This book is a revision of that dissertation, which I was fortunate enough to write under the supervision of Thomas Nagel. My debt to him will be obvious to all those who know his work, and even more obvious to those who know him personally. If I have any sense at all of philosophical depth, it is thanks to Tom; if I do not, it is not for lack of his trying to teach me.

After finishing the dissertation, several years were pleurably consumed while I tried to find answers to objections and criticisms gleefully thrown my way by friends and colleagues; looking back, it seems to me that Kurt Baier, David Gauthier, Tim Scanlon, and Michael Slote were among the most gleeful. I think the book got better as a result of this process; at any rate, it certainly got longer.

It got longer still thanks to the extraordinary and painstaking attention showered on it by Derek Parfit. Derek commented on the whole, not once, but three times, and I have incorporated his suggestions in well over a hundred passages. My thanks go beyond this, however, for his general encouragement and enthusiasm have themselves been equally important to me.

Years ago, in graduate school, my friends and I would console ourselves (for not having published anything) by reciting the motto 'Good times and high standards'. That my standards have not been high enough will soon, I fear, be all too apparent to the reader; but that the times have indeed been good I can readily attest, and to all those who helped make them so, I am grateful.

This page intentionally left blank

1 Against Ordinary Morality

Dramatis Personae

Morality requires that you perform—of those acts not otherwise forbidden—that act which can be reasonably expected to lead to the best consequences overall.

Few of us believe this claim; none of us act in accordance with it. Consider just how radically demanding it is. It bids us to act not with an eye to merely furthering our own projects and interests, or those of some individuals we may favor—but with regard for the interests of all individuals, the world as a whole, overall good. It demands that I ask how I can make my greatest possible contribution, all things considered—even though this may impose considerable hardship on me—and it forbids me to do anything less. If the claim is correct, most of my actions are *immoral*, for almost *nothing* that I do makes optimal use of my time and resources; if I am honest with myself I will recognize that I constantly fail to do as much good as I am able.

When I go to the movies I may spend a few dollars and enjoy myself for an hour or two. The pleasure I get is genuine, and it seems absurd to say that I have done anything *wrong*. Yet this is exactly what the claim entails, for both my time and my money could be better spent: the pleasure one could bring in an evening visiting the elderly or the sick quite outweighs the mild entertainment I find in the movies; and the money itself would have done much more good were it sent to famine relief—for even a few dollars is sufficient to enable another human being to survive a temporary food shortage brought on by drought. If the claim is right, then in going to the movies I do what is morally *forbidden*. This strikes us as wildly implausible; we agree that it would be *meritorious* to visit the elderly and donate the money to charity—but no more than that. If the claim is right, however, it is not merely

that it would be nice of me to forgo my slight pleasure: it is morally *required*.¹

To live in accordance with such demands would drastically alter my life. In a sense, neither my time, nor my goods, nor my plans would be my own. On this view, the demands of morality pervade every aspect and moment of our lives—and we all fail to meet its standards. This is why I suggested that few of us believe the claim, and that none of us live in accordance with it. It strikes us as outrageously extreme in its demands—so much so that I shall call its defender the *extremist*. The claim is deeply counterintuitive. But it is true.

This book is part of a defense of the extremist's position. Yet little will be said supporting it directly. For the most part, I want to critically examine the alternative position which is almost universally believed. If the difficulties with this alternative can be shown, then perhaps the extremist will be taken more seriously.

The alternative position is so widespread (in our culture, at the very least) that we can refer to it as *ordinary morality*. Although there is disagreement over the details of this position, certain broad features are shared by the different versions. I have already pointed to one of these features—the claim that there is a limit to what morality requires of us. Many acts which would lead to the best results overall² nonetheless are *not* required of us by ordinary morality, typically because the sacrifice would be too great to demand it of us. I am not required to devote my free time to fighting political oppression, nor must I give up my luxuries to support cancer research.

¹ I use the term 'required' as the positive counterpart of 'forbidden'—i.e., a given act or (act-type) is required if and only if not performing that act (or act-type) is forbidden. I use this somewhat unusual term in the hope of avoiding fruitless disputes over the proper scope of more traditional terms such as 'duty' or 'obligation'; for on some accounts of these terms, even though performing an act is neither obligatory nor a duty, it might still be forbidden to fail to perform that act. Similarly, on some accounts, 'ought' is too weak a term for my purposes: it may be that one ought to perform some act even though not performing the act is not forbidden. I should also note that I reserve the terms 'forbidden' and 'required' for moral judgments all things considered; that is, as I understand these terms it is improper to say, e.g., 'Fritz is required to do X, but all things considered he ought to do Y'; and so on. Finally, as I intend to use the term 'wrong' (departing from some accounts), if an act is wrong it is forbidden.

² The extremist's position presupposes the availability of an agent-neutral ranking of possible outcomes according to overall value. Although the notion of such a ranking is controversial in other contexts, it won't be examined here—for the dispute between the defender of ordinary morality and the extremist is not over whether we can *rank* possible outcomes, but rather over whether we are required to *pursue* that outcome we judge to be best from a neutral perspective. I will also leave unexamined the question of whether there is a uniquely correct ranking of worlds, and such familiar problems as decision-making under uncertainty and whether this requires a cardinal (and not merely ordinal) ranking, whether the extremist should require maximization of expected value of outcomes, what sorts of ignorance and miscalculation on the part of the agent will be excused, and so on. These problems are important; but not for the issues which concern us in this work.

On this view morality permits me to favor those things I most care about—whether my own welfare, the welfare of others, or any projects to which I am committed. It will be convenient to have a single term covering all of those things in which I take a special interest and which, as a result, I may want to favor. Let us call all of these things my *interests*, recognizing that this use of the term goes somewhat beyond its ordinary scope. In particular, it should be noted that my interests, as I will use the term, need not be limited to what is in my self-interest as that is normally understood. Rather, my interests include all of the objects of my concern, and so may include the well-being of family or friends, as well as various impersonal goals that I support, which may have little or no connection to my own individual welfare.³

On the view of ordinary morality, then, I am permitted to favor my interests, even if by doing so I fail to perform the act which leads to the best consequences overall. Since the agent is given the option of performing (or not performing) acts which from a neutral perspective are less than optimal, we may call such permissions *agent-centered options*, or more briefly, *options*. The first feature of ordinary morality, then, is the belief in the existence of options.

This is not to say that no sacrifices at all are required by ordinary morality, but they *tend* to be rather modest and limited. If a child is drowning in front of me, and I can save her by throwing a life preserver, ordinary morality certainly requires me to do so—even though it takes some slight physical effort, and my clothes may

³ In many contexts, it is a controversial philosophical question to what extent an individual's interests (in my broad, inclusive sense) rather than her *self*-interest (more narrowly understood) is the appropriate object of concern (whether on the part of the individual herself, or on the part of others). Given the inclusive meaning I have just adopted, my regular use of the term 'interests' will abandon neutrality on several aspects of this issue. I will, at a few relevant points, have something to say in defense of this practice, but these remarks will hardly be decisive. Accordingly, some readers may find certain portions of the discussion more plausible if (when appropriate) they interpret my talk of 'interests' as referring only to 'self-interest'.

become soaked. Generally, however, I am not required to significantly sacrifice my interests in order to provide aid to others—even though objectively greater good would result from my doing so. Of course I am *free* to make such sacrifices if I choose to—and morality encourages me to do so—but these acts are not *required* of me: there are agent-centered options.

The second broad feature of ordinary morality is that it lays down certain strict limits on our actions—forbidding various types of acts *even* if the best consequences overall could be achieved only by performing such an act. I may not murder my rich uncle Albert in order to inherit his wealth. Nor may I murder him even if this is the only way to guarantee that his millions get spent on famine relief—my act thus saving many more lives than it takes. Restrictions arising from rights, prohibitions against intending harm, and the like, are all examples of this second feature. A second class of examples of this second feature can be found in the special obligations to others that can arise due to past promises, or through institutionally defined roles, such as family or professional duties, and so on. Here too, ordinary morality forbids violating such special obligations, even if this is the only way to achieve the best consequences overall. Advocates of ordinary morality may differ as to whether these various restrictions and special obligations are absolute, or may be violated in pressing enough circumstances—but all are agreed that the limitations they impose cannot normally be transgressed. Since both the restrictions and the special obligations forbid agents to perform acts which from a neutral perspective might be optimal, let us call such limits *agent-centered constraints*, or more briefly, *constraints*. The second feature of ordinary morality, then, is the belief in the existence of constraints.

Both features of ordinary morality temper the relentless pursuit of the best consequences. Options give the agent permission to pursue his own interests rather than the overall good; constraints forbid certain courses of action—even when they are necessary means to achieving the overall good. Within the limits of the constraints, there *are* occasional demands, but typically only rather modest ones are made.

It is clear that ordinary morality stakes out a more moderate position than the one espoused by the extremist. It imposes certain limits, and requires certain sacrifices—but it does not come near, as does the extremist's morality, to pervading and dictating every

aspect of our life. For this reason, I shall call the defender of ordinary morality the *moderate*.

As I have noted, on some views of ordinary morality behavior which is normally forbidden by a constraint may be permissible under pressing enough circumstances (e.g., when the number of lives at stake is sufficiently great). If there is such a cutoff point, beyond which the constraint is ‘relaxed’, we may say that the constraint has a (finite) *threshold*. Along a similar vein, it is also worth remarking that options need not be unlimited either—but may instead have thresholds beyond which the agent is not permitted to react in the normally permissible manner. Now in principle, such thresholds (whether for constraints or options) could be extremely low. If this were the case, the corresponding constraints and options would temper the pursuit of the good in only a very limited way: whenever much of anything was at stake (in terms of the overall good) the thresholds would be crossed, and the agent would be required to promote the greater good. Thus the mere presence of constraints and options does not in itself logically guarantee that the resulting position is significantly less extreme than that of the extremist.

In fact, however, if a view is to capture anything close to our ordinary moral beliefs, thresholds will generally have to be quite high (if not infinite). And if this is the case, the presence of constraints and options will indeed significantly temper the relentless pursuit of the overall good, yielding a much more moderate position. As the defender of ordinary morality, then, the moderate deserves his name.

Our list of *dramatis personae* will be complete with the introduction of the *minimalist*. His position will be less sharply defined, for in this essay he will act largely as a foil to the moderate. His purpose is to remind us that the moderate is trying to maintain a middle position, and as such can perhaps be criticized for having gone too *far* in his demands. Minimalists, of course, may differ with one another as to which of the moderate's demands are to be rejected; but for our purposes there will be no need to offer a more fine-grained classification.⁴ Provided that one can coherently

⁴ Note that a variety of sharply diverging positions will fall within the minimalist camp, including egoists (who believe that one is never required to sacrifice overall self-interest), nihilists (who believe that everything is morally permitted), and extreme libertarians (who recognize the validity of constraints, but deny that there is a moral requirement to provide aid).

imagine *some* position rejecting the given demand on the grounds that far less is actually required than the moderate claims, I will offer such a rejection as the view of the minimalist. For it is important to bear in mind that ordinary morality can indeed be challenged as too demanding, despite its being a more moderate position than that of the extremist. Thus, as we have noted, the moderate does require some minor sacrifices—e.g., to save the drowning child. The minimalist, however, argues that this is already to ask too much: one is never required to aid another. Similarly, the moderate rules out certain courses of action—e.g., murdering my uncle for personal profit—but some minimalists deny that such acts are prohibited.

It should be clear that all of the positions we have marked out encompass broad ranges of more specific views which have important features in common. Reducing the multitude of possibilities to three broad types,⁵ however, allows us to focus on a central philosophical problem: the moderate is under simultaneous attack from both sides. He must defend his position from both the extremist *and* the minimalist—or admit defeat. The tension inherent in his position is apparent. By abandoning the ideological purity of the two ends, the moderate risks internal incoherence. This is not to say, of course, that the mere fact that the moderate lies in the middle is in itself reason to assume that his position is implausible. But if, as I believe, the moderate's view cannot be given a coherent justification, its vulnerability to attack from both sides may make this easier to see. Forcing the moderate to defend his position against the extremist's demands without collapsing into the arms of the minimalist may expose that lack of coherence.

As we have seen, the moderate denies that we are morally required to do all that we can to promote the overall good. In contrast, the extremist claims that in fact there is such a requirement. It should be pointed out, however, that in making this claim, the extremist is not committing himself to any particular *account* of the good. That is, the extremist asserts that we must each make our greatest possible contribution to the overall good, but this claim does not in itself specify which factor or factors make one outcome better than another—i.e., what it is by virtue of which

⁵ Even these three do not exhaust the field, but there is no room here to consider all the possibilities. In Chapter 5, however, we shall examine the position of the neo-moderate, who embraces options while rejecting constraints.

one outcome is better than another. Individual extremists may well differ on this question, as may moderates. Presumably, on all plausible accounts of the good, human well-being will be a central component. (For this reason, my examples will involve this crucial factor.) But it should be noted that nothing prevents the extremist from adopting a *pluralist* theory of the good—i.e., one that gives independent weight to several factors. For example, it might be claimed that various distributional factors are relevant to the goodness of outcomes:⁶ all things being equal, one outcome may be better than another if well-being (or, more broadly, the satisfaction of interests) is distributed according to effort, or desert, or in an egalitarian manner. And other factors might be brought in as well.

Articulating and defending an adequate theory of the good is an important task for moral philosophy, but not one that will be addressed in this work. The point to bear in mind is simply this: whatever the most plausible theory of the overall good, it can be incorporated into an extremist framework. Thus the extremist and the moderate need not be conceived as disagreeing over the correct account of the good. Rather, the point of disagreement is simply whether or not we are morally required to do all that we can to *promote* the overall good.

The moderate might be tempted to object that an agent can go overboard on sacrifices—exhausting and impoverishing herself—ultimately destroying her ability to make continual contributions to the overall good (however this is ultimately understood). In this way the extremist's position might be thought to be self-defeating: were it only more moderate in its demands, agents would—in the long run—be able to do the world more good.

Such an objection, however, would rest on a misunderstanding of the extremist, who does not demand sacrifices for their own sake—but only insofar as they are the cost of producing the greatest possible good. The extremist would be the first to urge that mindlessly driving oneself to exhaustion or recklessly dispensing one's goods can be counterproductive. Far from being required, such behavior is forbidden. What each agent is required to do is to

⁶ Since, on a pluralist theory of the good, the very performance of a given type of act may itself be a factor in how well the history of the world goes, talk in the text of the goodness of the 'outcome' of a particular act should not be understood in a narrow sense, i.e., limited to what causally follows from the act, or what happens after the act. (Similarly for talk of 'consequences', 'results', and the like.)

act in such a way that she can make her greatest possible contribution to the overall good (given her own particular talents). Very likely this involves taking a hard look at her life plans, and reshaping them accordingly; at the very least it involves taking into account the long term effects of her action and not only the more immediate ones—and this is quite enough to explain the need for relaxation and a judicious apportionment of resources.

There *is*, however, an important qualification in the extremist's claim which we have not yet considered. According to the extremist, morality does not straightforwardly require you to choose that act, *whatever* it is, which can be reasonably expected to lead to the best consequences. Rather, it restricts your choice to those acts not otherwise morally forbidden. However, the extremist's claim itself is neutral on the issue of whether morality ever *does* forbid an act which would lead to the best consequences. Let us consider both possibilities.

One of the central concerns of this book will be to examine whether agent-centered constraints—which create barriers to promoting the good—can be justified. For the moment, suppose that they cannot be. In this case, the qualification drops out as vacuous, and the extremist requires that you perform the optimal act, *simpliciter*. This is consequentialism.

Perhaps, however, constraints *can* be justified. Such a result is still compatible with the extremist's claim, and the extremist will continue to demand far more of an agent than the moderate. I may be forbidden to murder uncle Albert to free his vast fortune for famine relief, but nonetheless I am permitted to donate my own more modest fortune, and if the extremist is right I am required to do so. If there are constraints, then you are required to perform the optimal act among those acts which do not violate them.

It might be thought, however, that the extremist will not provide a significant alternative to ordinary morality if he recognizes the existence of special obligations. But this is a mistake. Consider, for example, an obligation to provide my children with a decent education. Admittedly, meeting this obligation may require the expenditure of a portion of my resources which could do even *more* good were it devoted instead to famine relief. Such obligations are, after all, constraints, providing potential barriers to the performance of optimal acts. But it would not be plausible to claim—and

certainly the defender of ordinary morality would not want to claim—that *all* of my resources are tied up in this way through special obligations. Typically, the most significant portion remains unconstrained. Of course there may be exceptions—rare cases where an individual's special obligations are *quite* demanding, and make an exhaustive claim on his resources. But such cases are, indeed, exceptions. For most individuals the special obligations recognized by ordinary morality are modest enough to leave the individual with a considerable portion of his resources. It is here that the moderate believes that options come into play: provided that I have met my obligations to my family (and any other special obligations I may have), I am permitted to spend my remaining time and money as I choose. But if the extremist is correct, then once my special obligations are met I am required to devote myself to making my greatest possible contribution to the overall good. Presumably this rules out doing many things that the moderate believes to be permitted—favoring my family even more, say, with expensive gifts. Therefore, even if the extremist recognizes the existence of constraints, including special obligations, by denying the existence of options he provides a significantly more demanding position than ordinary morality.

I will often express this fundamental difference between the extremist and the moderate by saying that the extremist accepts, and the moderate rejects, a *general* requirement to promote the good. A few words of explanation about this expression may be helpful. After all, even the moderate believes in *occasional* requirements to promote the good, since in the right situations (e.g., the drowning child) the agent may well be required to act in a way that will lead to the greatest good overall. What the moderate denies, however, is the claim that the agent must—in *general*—do all he (permissibly) can to promote the good. The moderate believes in the existence of options, and these will often permit the agent to act in a way that fails to make his greatest possible contribution to the overall good. It is, then, to the moderate's embrace of options that I am pointing when I say that the moderate rejects a general requirement to promote the good. Second, as we have just seen, if there are indeed constraints, then even the extremist will not accept an *unqualified* requirement to promote the good. What the extremist is committed to, rather, is a requirement to promote the

good—*within* the limits of constraints (if there are any). When I say that the extremist accepts a general requirement to promote the good, this (often implicit) qualification should be borne in mind.

Properly understood, then, the extremist's claim is itself neutral on the issue of whether or not there are any constraints. Either way, the extremist goes beyond ordinary morality by insisting on a general requirement to promote the good. The extremist is perfectly capable of recognizing the existence of constraints—if such there be—while still denying the existence of options.

As we shall see, however, the most straightforward extremist position recognizes neither options *nor* constraints, while the moderate's *defense* of options *requires* him to establish the existence of constraints. For this reason I shall portray the extremist as dubious of the claim that there are constraints, and shall place the burden of establishing their existence squarely upon the shoulders of the moderate. But, as we have just noted, even if there are constraints the extremist offers a significant alternative to the moderate—a far more demanding alternative.⁷ In opposition to it, the moderate insists upon the existence of options: admirable as it may be to sacrifice one's career, time, possessions, or life for the greater good, such behavior—he assures us—is surely not *required* by morality. As long as one stays within the bounds set by moral constraints, and makes the typically *minor* sacrifices occasionally required, one does all that morality demands of us. To the minimalist, however, even the demands of the moderate are unreasonable: a minor sacrifice is a sacrifice for all that, and there is no justification for requiring that an individual forsake his interests for the greater good.

The moderate stands squeezed in between the extremist and the minimalist—and I intend to increase the pressure. We are going to ask the moderate to try to justify ordinary morality. Yet the moderate may feel secure; after all, ordinary morality seems to be supported by our intuitions. Isn't this an adequate justification? I believe, however, that intuitive support is *not* adequate for justifying a moral theory, and it is important to see that this is so. Let us take a brief look at one aspect of the method of moral philosophy.

⁷ Interestingly enough, it seems that few extremists have actually embraced this position. I believe that Godwin can be interpreted as such an extremist—one who denies options, but recognizes constraints—but he is the only prominent defender of this view of which I am aware.

Moral Methodology

Let me start with a model of moral theorizing which is, I believe, inadequate. On this account we begin moral philosophy with a set of pretheoretical moral intuitions—beliefs about the moral character of a variety of specific situations (both actual and hypothetical). Our goal—on this model—is to discover a set of moral principles which will yield judgments about the range of possible cases, striving to discover those principles which give the closest possible ‘fit’ to our original intuitions—which act as our data.

But not just any principles will do on this account—for principles of *one* sort could *always* be cooked up to fit our intuitions exactly. If this were sufficient, we would need to do no more than translate those intuitions directly into *ad hoc* lists: in situations of type 1 do A; of type 2 do B; of type 3 do C. . . Such *ad hoc* principles are unsatisfactory, for we want our moral principles to amount to more than shopping lists. In moral philosophy we want to apply the same sort of criteria that we use for theory building quite generally: we want our moral theory to have simplicity, power, and coherence.

There is no room here to give these matters the attention they deserve, but *roughly*, a moral theory has *simplicity* if it yields a body of judgments out of a relatively sparse amount of theory, deriving the numerous complex variations of the phenomena from a smaller number of basic principles. The shopping lists fail on this criterion, requiring a principle for each case. A moral theory has *power* when it yields judgments not included in the original data base. The shopping lists fail here, too, being no help in those cases where our original intuitions offer us no guidance. *Coherence* on this account can be taken as consistency, and if we find ourselves of two minds for some cases—with incompatible stances both having some intuitive plausibility—then the shopping list approach will fail once more. Rejecting the shopping lists, then, the model is only satisfied with principles which have simplicity, power and coherence.

Finding principles to fit out intuitions which are acceptable by these criteria is not a trivial matter: generally we can expect that the fit will not be perfect—and the principles we arrive at will differ from our initial intuitions for some cases. Hopefully, however, the principles themselves will have some intuitive plausibility—and at the very least their ability to yield judgments with which we are substantially in accord gives them some credibility. Thus, the

principles themselves may give us grounds for rejecting some of our initial intuitions, and the altered set of intuitions may in turn give us grounds for altering the principles. In this manner, we work back and forth between intuitions and principles until a stable point is reached. The remaining intuitions represent our considered judgments, and the moral principles at which we have arrived can guide us in difficult cases. The entire process forms a justification of the end products: the underlying principles justify our intuitions, and the intuitions in turn support our claims for the underlying principles.⁸

Now in several ways this model is misleading. It is doubtful that we have any *pretheoretical* intuitions, and so our actual intuitions cannot provide a special foundation for theorizing. It would be more correct to say that we begin moral philosophy already possessing some moral theory—albeit theory which is only half-formed, and largely unarticulated. We have a variety of beliefs—about specific cases, and about general moral principles. We are more wed to some of these than to others; there are those we are more immediately disposed to assent to while others require reflection; and so on. As we critically reflect, and face new experiences, our theory evolves, striving for greater simplicity, power, and coherence. We abandon beliefs, adopt new principles, and our intuitions may change—conditioned by our altered judgments. If we ever reach a stable point, it may be far removed from the point where we began.

The original account is even more misleading in what it does *not* say, for it fails to bring out adequately the importance of *coherence*. It is not sufficient justification for a set of moral principles—even when they have simplicity, power, and are mutually consistent—that they yield (the bulk of) our intuitions about specific cases. All that this guarantees is that we will have succeeded in axiomatizing the moral view, so to speak. But we want more from a moral *theory*. The original model would in effect be satisfied with a set of maxims, guiding our behavior. For it, coherence is only a matter of consistency: the maxims must not give contradictory advice. But surely there is more to coherence than this. We want the principles to hang together, to be mutually supportive, to be jointly illuminated by the moral concepts to which we appeal. If the

⁸ Readers will doubtless recognize some semblance of Rawls' notion of reflective equilibrium in the hasty sketch I've given here. Cf. Rawls, pp. 19–21, 46–51.

principles are only maxims, albeit consistent maxims, then they are not enough. Rather, they must form, or be part of, a coherent moral theory.

Seeing that the criterion of coherence is richer than that of mere consistency should also lead us to recognize the importance of a fourth criterion for moral theories—one that seems altogether neglected by the original model: a theory must be *explanatory*. As noted, it seems that the original model would be satisfied with a mere set of maxims. But even if the maxims form a mutually consistent set, we *still* want our theory to provide an *account* of the distinctions, goals, restrictions, and the like, which they embody as well. An adequate justification for a set of principles requires an *explanation* of those principles—an explanation of why exactly these goals, restrictions, and so on, should be given weight, and not others. Short of this, the principles will not be free of the taint of arbitrariness which led us to move beyond our original *ad hoc* shopping lists. To reduce the many to the few is not yet to *support* the few. Unless we can offer a coherent explanation of our principles (or show that they need no further justification), we cannot consider them justified, and we may have reason to reject them.

This need for explanation in moral theory cannot be overemphasized. We want our moral theory to help us to understand the moral realm. (Indeed, one of the things we want our moral theory to help us to understand is how there can even be a moral realm, and what sort of objective status it has.) A large part of the motivation for the criteria of simplicity and power is the drive to understand: by reducing the complex to the simple, we can hope to understand the complex; by having a powerful theory we can hope to move from a base which is understood to an understanding of new areas. Power, simplicity, and consistency are valuable in themselves, but it is the need to understand the moral realm which is, I believe, paramount. Ultimately, unless we have a coherent explanation of our moral principles, we don't have satisfactory ground for believing them to be true.

Suppose that in analyzing our intuitions, we find a distinction which if appealed to yields our intuitive judgments. Morally, we still need to know whether the difference *ought* to make a difference. Perhaps a slaveholder might find that a principle which distinguished according to skin color yielded intuitively correct

judgments about when a gentleman is morally required to aid someone being whipped, and when he is not. Merely having *found* the distinction underlying his intuitions is not sufficient to *justify* it. We want to know *why* difference in skin color should support differential treatment. If the slaveholder cannot offer an explanation, then the distinction hangs free of the rest of his moral theory, and considerations of coherence give him reason to reject the distinction as morally irrelevant, as well as repudiating the intuitions which turn on it. We might call such cases where there is a critical lack of support *dangling distinctions*. An adequate defense of a moral theory must not leave its distinctions dangling.⁹

Most commonly, of course, a distinction will not be left *totally* unsupported. Generally, *something* can be mustered in its defense. The key issue is really whether the support offered will withstand critical examination; and the criticism may well be internal. Our slaveholder may find that other principles to which he is committed militate against the moral relevance of the black/white distinction. And presumably, at the very least, he is opposed to *others* drawing lines they cannot support—and he may turn this standard back upon himself. In such cases, there is an internal tension in the moral theory, pushing it toward revision. We cannot say in advance how the tension will be best resolved—two or more directions may suggest themselves—but it is especially this drive to explanatory coherence which may produce a moral theory quite unlike the original.

It may be objected that explanations have to come to an end *somewhere*. Perhaps this is so, but it would still be no license to cut off explanation at a superficial level. If a distinction stands isolated, or is at odds with more firmly supported beliefs, we have grounds for rejecting it, despite its intuitive appeal. What I want to suggest is that the distinctions which underlie the moderate's position dangle, and appeal to their intuitive support is no more adequate than such an appeal would be for the slaveholder.

Given the nature of the internal tension in ordinary morality, there are two directions moral theory can go: toward the extremist,

⁹ It may be that Rawls intends reflective equilibrium to encompass the elimination of dangling distinctions. Or maybe not. I won't argue the point of exegesis here. (It may also be worth noting that the methodology of some so-called 'intuitionists' may not differ substantially from the one I am endorsing here. Arguably, Sidgwick is among this group; but once again, I'll not pursue the point of exegesis.)

or toward the minimalist. My sympathies—as I've indicated—lie with the extremist, but I'll not attempt criticism of the minimalist here. To attack ordinary morality is ambitious enough.

We should be clear about just what such an approach can accomplish. Even if the internal incoherence of the moderate's position can be shown, it will not necessarily make the intuitions which support it disappear. Even if the slaveholder is persuaded to abandon his racist theory, he may continue to intuitively *feel* that there is a difference. But he will reject such intuitions as unjustified—a sort of ineliminable moral illusion, similar to certain optical illusions which do not lose their intuitive hold on us even when our theory tells us better.¹⁰

This freedom of moral philosophy to reject many of our intuitions is of a piece with its ability to justify moral theories radically different from our original beliefs. In criticizing the moderate, I am *not* merely arguing that although ordinary morality must be given a different foundation than we generally believe, the moderate is, nonetheless, basically correct about moral practice.¹¹ Nothing so comforting. Rather, I claim that our moral beliefs and practices need to be drastically revised.

Promoting the Good

The moderate may believe that there is a quick path to justifying ordinary morality. He may be tempted to argue that ultimately one only has reason to do what one is motivated to do. (This needs to be made a bit more sophisticated, but it will do for now.) It is then a simple psychological fact that most of us are disposed to help those suffering in front of us when it is not difficult to do so, and not otherwise; and that most of us are disposed to refrain from hurting others so long as they refrain from hurting us. The moderate may suggest that this is all the justification that ordinary morality needs.

¹⁰ A complete explanation, of course, will also have to *account* for the rejected intuitions.

¹¹ This seems to be the opinion of many consequentialists. Mill, e.g., claims that generally one need only make sure that one's act violates no one's rights, for 'the occasions on which any person (except one in a thousand) has it in his power . . . to be a public benefactor . . . are but exceptional; and on these occasions alone is he called upon to consider public utility; in every other case, private utility, the interest or happiness of some few persons is all that he has to attend to' (Chapter 2, paragraph 19).

Such an approach, however, would leave it too contingent a matter whether an agent has reason to refrain from killing others: if the slaveholder has no feelings of sympathy for his slaves, he may have no moral reason not to kill one who displeases him. And so on. This is an unacceptably minimalist conclusion for most moderates. Therefore, they must forgo the shortcut, and attempt a more substantial justification.¹²

The need to provide such a justification for ordinary morality will be felt even more acutely when we realize that the moderate himself recognizes a reason, all things being equal, to promote the greater good. It will therefore be all the more pressing for him to justify his belief in constraints and options which temper the pursuit of the good.

Suppose a building is on fire. Upon entering, I find a child and a bird trapped within. Needing one hand free to clear a path back outside, I can only save one of the two, and I hastily pick up—and escape with—the caged bird.

Clearly I have done something wrong. Even if the moderate believes that I was not morally required to risk my safety by entering the building in the first place, he nonetheless believes that once I have decided to undertake the risk, I should have promoted the greater good, by saving the child. If my interests are equally affected by either of two courses of action, I have reason to pick that act with the objectively better outcome.

The moderate even believes that there are cases where I am morally required to act so as to promote the good even though there is some cost to myself which I am not inclined to take on: e.g., I am required to throw the drowning child the life preserver, even though my clothes will get soaked, and I risk catching a cold.

If the moderate is asked to account for his judgments here, the best explanation seems to involve the quite general thesis that one

¹² It may be objected that my reply presupposes that moral principles are universal—binding on all agents. But this is not so. It only assumes that advocates of ordinary morality would be unwilling to let typical agents escape so easily from, e.g., a requirement not to kill. At any rate, even those moderates who see themselves as doing no more than describing their personal (or societal) moral code should, I believe, want that code to have the sort of internal coherence I described in the previous section. We want to be able to offer a satisfying explanation of the various aspects of our moral code; it should be coherent—at least by our own lights. The problem of dangling distinctions should be felt even by those who make no claim to universal (or nearly universal) moral principles; and so the shortcut is not sufficient. (We'll return to the issue of universality in Chapter 8.)

always has a (morally acceptable) reason to promote the good. Now this thesis should not be misunderstood as being more bold than it is. It does not claim that one always has an *overriding* reason to promote the good; there may well be morally acceptable countervailing reasons at times, whose net effect is to outweigh the standing reason to promote the good. Presumably the moderate believes that there frequently *are* such countervailing reasons: some merely make it *permissible* to refrain from promoting the good; others may *require* it. But, although the reason to promote the good may be overridden, it does not disappear: it is a *pro tanto* reason.

The term ‘pro tanto’ may be somewhat unfamiliar (partially, perhaps, because my use of it is also slightly ungrammatical), so a word of explanation may be in order. A *pro tanto* reason has genuine weight, but nonetheless may be outweighed by other considerations. Thus, calling a reason a pro tanto reason is to be distinguished from calling it a *prima facie* reason, which I take to involve an epistemological qualification: a *prima facie* reason *appears* to be a reason, but may actually not be a reason at all, or may not have weight in all cases it appears to. In contrast, a pro tanto reason is a genuine reason—with actual weight—but it may not be a *decisive* one in various cases.¹³ The claim, then, is that the best explanation of the various judgments that the moderate wants to make involves the acceptance of a standing, pro tanto reason to promote the good.

The belief in this pro tanto reason is common ground for the moderate *and* the extremist. The moderate proceeds to argue for the existence of other, potentially overriding, reasons as well. But in the most straightforward case, the extremist simply *stops* with this common belief. He shares with the moderate the belief that promoting the good is a reason for performing an act; and he simply recognizes no other morally acceptable reasons at all. This makes more urgent the moderate's need to find a justification for his belief in constraints and options. For he himself recognizes the grounds for the extremist's position; he cannot deny the pro tanto reason to promote the good without retreating into the minimalist's

¹³ It may be helpful to note explicitly that in distinguishing between pro tanto reasons and prima facie reasons I depart from the unfortunate terminology proposed by Ross, which has invited confusion and misunderstanding. I take it that—despite his misleading label—it is actually pro tanto reasons that Ross has in mind in his discussion of what he calls prima facie duties. See, e.g., Ross, pp. 19–20, 28–29.

camp (although there would still, of course, be some disagreements between the moderate and the minimalist, e.g., over the existence of constraints). Indeed, to the extent that options and constraints are at ideological odds with the drive to promote the good, we may well wonder whether the moderate will be able to provide a philosophical house capable of holding all three.

Of course, it should go without saying that the belief in the existence of a *pro tanto* reason to promote the good is itself in need of defense. After all, the claim that there is a standing reason to promote the overall good is one that will be rejected by the minimalist—who will not be satisfied with the observation that acceptance of such a reason is compatible with the belief that it is frequently overridden. For the minimalist sees no reason to accept the *pro tanto* reason to promote the good in the first place. Obviously enough, the mere fact that belief in such a reason is common to both the moderate and the extremist hardly constitutes an adequate justification for that belief.

Examining whether there is indeed any kind of reason at all to perform an act simply by virtue of the fact that it will lead to a greater amount of good overall is, then, yet another essential task for moral philosophy. A defense of the *pro tanto* reason to promote the good would be an important element in a complete justification of ordinary morality, and, as I have just suggested, it would be the crucial step in a defense of the most straightforward extremist position. Nonetheless, in this work I will not be considering whether the *pro tanto* reason to promote the good can, in fact, be successfully defended. As I have indicated, my primary, driving interest in this work is to see whether the moderate can defend ordinary morality against the greater demands of the extremist (without, of course, retreating into what the moderate takes to be the inadequate moral demands of the minimalist). Since the belief in the *pro tanto* reason to promote the good is common to the moderate and the extremist, it is not an element of ordinary morality that the extremist will want to challenge. Indeed, as we have seen, by emphasizing the moderate's acceptance of that reason, the need for an adequate defense of options and constraints is made all the more pressing.

There may, however, be moderates who will resist my suggestion that they are committed to the existence of a *pro tanto* reason to promote the good. I certainly do not mean to claim that this is an

explicit component of all versions of ordinary morality. Rather, my claim is that the existence of such a reason provides the best explanation of various judgments that the moderate wants to make; and so, wittingly or not, the moderate is implicitly committed to such a reason. I will defend this claim in greater detail in Chapter 2. My purpose here is simply to bring out how the recognition or acceptance of such a reason sharpens the bite of the general demand that the moderate provide an adequate justification for ordinary morality. But it is worth stressing that the need for such a justification was already established in the previous discussion of moral methodology. Even if, for the time being, some moderates resist the *pro tanto* reason to promote the good, I have argued that they must recognize the need to provide a more adequate defense of constraints and options than the mere appeal to intuitions.

Resisting Options to do Harm

How then is the moderate to justify his belief in constraints and options? Let us start by observing the logical independence of these two features of ordinary morality: a moral theory needn't have both (nor either, of course). A system might have options without constraints; an extreme case of this would be the minimalist system in which the agent is permitted to do absolutely everything—there being no requirements at all. Similarly, a system might have constraints without options; we mentioned such a possibility when noting that the extremist could concede the existence of constraints and still insist that the agent is required to do the optimal act within the limits of those constraints. Thus, a defense of one feature might be able to proceed without defending the other. Since the dispute between the extremist and the moderate turns on the existence of *options*, let us consider how the moderate might go about defending them.

For the extremist, the requirement to pursue the good pervades an agent's entire life—all its aspects, every moment. The moderate wants to carve out a sphere in which the agent can find relief from the demands of morality which would otherwise be constantly pressing upon him. Options provide such a protected sphere, and within it agents are free to pursue their own interests, at the possible expense of the good.

The problem facing the moderate is this: a moral system might include not only options to *allow* harm, but also options to *do* harm.¹⁴ For example, a system with no requirements at all includes an option to bring about harm no matter how bad the outcome. I am not saying that such an option is particularly *plausible*—but its theoretical possibility must not be neglected. The moderate might respond that what was desired was that the demands of morality should be limited—not eliminated. But this reply won't really help, for we can always make the option to harm more modest, without yet making it acceptable. If the option to harm is given a finite threshold, or is restricted to particular forms of harm—say, harpooning newborn males—then we will have succeeded in giving the agent a protected sphere, without giving him complete license. This point is worth stressing: moral systems can provide options of a sort which are unacceptable to the moderate.

Strictly speaking, of course, it must be admitted that not all options to do harm are unacceptable to the moderate. He may well believe, for example, in an option to kill in self-defense, even when it is not optimal to do so.¹⁵ Such options are rare, however, and for simplicity of exposition, I will write as though the moderate *does* reject all options to do harm. (On certain broad, encompassing conceptions of 'harm', acceptable options to do harm will be somewhat more common; but this doesn't really affect the main point being made.) It is clear, at any rate, that many or most imaginable options to do harm would not be acceptable to the moderate. For example, we are not in general permitted to kill another for mere personal profit; the moderate would obviously find an option to do harm of this sort unacceptable. Yet if the motivation behind options is to provide the agent with a protected sphere, free from the demands of morality, it seems that such unacceptable options are just as likely to fit the bill as options of a kind that the moderate may find more palatable.

Thus we can see how, in fending off the attack of the extremist, the moderate is in danger of retreating into an overly minimalist position. Against the extremist's claim that agents are required to

¹⁴ In Chapter 3 I shall differentiate between the do/allow distinction and the intend/foresee distinction—but for the present, talk of doing harm and allowing harm is meant to be neutral between the two.

¹⁵ Permitting self-defense at all—even when it is optimal—may be problematic for some moderates; this point is discussed in Chapter 4.

do all they can to promote the greater good, the moderate asserts the existence of options. But the general motivation for options seems more powerful than the moderate may have realized: it runs the risk of leading to the construction of a minimalist system—a system that the moderate cannot embrace.

The moderate, of course, will have to support his case for options with more than a general plea that the agent should have some protected sphere. Perhaps the *specific* considerations which he offers in favor of options will support *only* options to allow harm, and *not* options to do harm. If this were so, then the problem would be avoided. But in fact the problem *cannot* be avoided in this way, for the most plausible and straightforward defense of options appears to support both options to allow harm *and* options to do harm. Let's see why.

If the extremist is right, there is no limit to what you might be called upon to sacrifice in the pursuit of the good. Your material possessions, time, effort, bodily parts, or life itself—all of these might be commandeered by morality, and put to purposes quite unlike those to which you would dedicate them were morality's demands less severe. Taking a stand against this view, the moderate seems to be the voice of reason itself when he claims that there surely must be a limit to the costs that morality can inflict upon an agent.

The most straightforward defense of options will be grounded in such an appeal to considerations of the cost to the agent which would be imposed by various moral demands. The greater the sacrifice which morality requires, obviously enough, the more significantly it will decrease the agent's ability to mold his life as he chooses and to promote his interests. The moderate may want to argue that more than a certain loss of such autonomy is morally intolerable. Hence, morality can only exact so much, and no more.

Obviously, a great deal more needs to be said about such an appeal to cost. (Hereafter, 'cost' means 'cost to the agent'.) This is only a sketch of an argument. Rather than pursuing these matters now, however, I want to put aside for the time being the question of whether an appeal to cost would in fact be sufficient to ground options,¹⁶ and ask instead: regardless of whether it is adequate to *justify* them, what *sorts* of options would an appeal to cost support? Now it seems intuitively clear that an appeal to cost will

¹⁶ We will return to this question in Chapter 7.

support an option to *allow* harm—for preventing harm can be quite costly. (It would be a limited option, of course, for I would still be required to act in those cases—e.g., the drowning child—where the cost of providing aid would be minimal.) The point I want to press here, however, is that an appeal to cost supports options to *do* harm as well as options to allow harm, and the moderate, therefore, may have to *abandon* his appeal to cost. For it is not only *preventing* harm which can be costly—*refraining* from harming can be costly too. And, if this is right, then an appeal to cost would support an option to do harm in such cases.

The claim that refraining from harming can be costly may not seem correct. After all, it takes no time, resources, or physical effort to keep from kicking a passing stranger whose face I happen to dislike. In general, it seems, refraining from performing some act requires no significant sacrifice. Thus *prima facie* requirements not to do certain sorts of acts seem immune to appeals to cost.

But not all requirements to refrain from harming exact only an insignificant sacrifice. Recall my wealthy uncle: suppose that by murdering dear old uncle Albert I stand to inherit one million dollars. The cost of complying with the requirement that I not kill is enormously high. It is not an out-of-pocket expense, to be sure—but we might speak of a *prospect* cost: I stand to gain one million dollars by killing my uncle, and if morality is going to close off that act to me, then it exacts a tremendous cost. If high cost is indeed ground for denying the existence of a requirement, then we seem to have such a case here—and it is permissible to kill uncle Albert. But this is quite unacceptable to the moderate.

The moderate might be tempted to argue that it is illegitimate to weigh the cost of not killing my uncle in at one million dollars. I have already noted that there is no out-of-pocket expense. Perhaps *prospect* costs are not sufficient to undermine the claim of a *prima facie* moral requirement.

On reflection, however, it is clear that the moderate can't make this move—for he does want to count prospect costs when it comes to *prima facie* requirements not to *allow* harm. Suppose a second rich uncle, Bruno, plans to leave me his million—unless I tell him to donate it instead to famine relief. Am I required to tell him to donate it? The moderate thinks not. Yet this is a case where I could prevent an immense amount of suffering, with little effort and no out-of-pocket expenses. Nonetheless the moderate wants to grant

me an option to allow the suffering and further my own interests instead. The reason, clearly enough, is the prospect cost: asking me to forgo the prospect of gaining one million dollars is asking too much.

So prospect costs *do* count, and if they tend to support options to allow harm in the Bruno case, they must also support options to do harm in the Albert case. But options to do harm are unacceptable to the moderate. Indeed, it is precisely those cases where the moderate most wants to insist upon the requirement not to kill—i.e., where the agent stands to gain the most by killing—that the appeal to cost tends to undermine ordinary morality.

Since the appeal to cost supports *both* sorts of options, if the moderate is going to maintain his defense of options to *allow* harm without being forced into accepting options to *do* harm, then he must argue that there are overriding *independent* reasons why an agent must not do harm. That is, he must offer an explanation of why doing harm is *forbidden*—an explanation backed by considerations forceful enough to override the pro tanto support *for* an option to do harm. He must defend the existence of a constraint against doing harm.¹⁷

If the moderate can first defend the existence of the appropriate sorts of constraints, then he may be able to argue as follows: ‘Considerations of cost tend to support the existence of options—both options to allow and options to do. I have already given strong reasons for thinking there is a constraint against doing harm—and these reasons override the considerations of cost insofar as the latter support some sort of option to *do* harm. But with regard to an option to *allow* harm, the considerations are not overridden. Consequently, I have provided a justification for an option to allow harm, to supplement my earlier justification for a constraint against doing harm.’

Despite the logical independence of constraints and options, then, the moderate's defense of the latter will require him to also establish the existence of the former as well. Only if the moderate can provide such an independent overriding defense of the constraint against doing harm, can he put forward the appeal to

¹⁷ Theoretically, the moderate *might* be able to provide reasons for not doing harm which were *not* strong enough to establish constraints against doing harm, but which *were* strong enough to override considerations of cost. This possibility will be examined in Chapter 5.

cost in opposition to the extremist, without fear that in doing so he will be backed into too minimalist a position. If he *cannot* provide a defense of constraints, then his defense of options will have to be abandoned.¹⁸

Constraints

Although we could discuss in a perfectly general fashion the difficulties of arguing for any sort of constraint, our primary interest in the question is motivated by the moderate's need to block various unacceptable options that would otherwise be supported by the appeal to cost. For example, the moderate needs to provide an overriding reason why I must not kill uncle Albert. Now cases involving the doing of harm may not be the only ones where the moderate would find an option unacceptable, but they are probably the most important such cases. In any event, if the moderate is to retain his use of the appeal to cost, then blocking or overriding that appeal in cases involving the doing of harm will certainly be essential to the moderate's defense of ordinary morality. So let us focus the discussion by examining the constraint against doing harm—and more specifically, the constraint against killing. (A discussion of constraints other than the constraint against harming would, of course, require corresponding alterations in the arguments about to be considered; in the case of certain special obligations, these alterations might be rather extensive. Still, the arguments given below are typical of those moderates offer in defense of constraints.)

In principle, our investigation could proceed by considering an example of what is doubtless the most common kind of case, i.e., one in which the best outcome will *not* be brought about by violating the constraint against killing. In such cases, of course, the

¹⁸ It may sound peculiar to speak of the moderate's *choosing* to abandon the appeal to cost should he find that he cannot resist options to do harm through the establishment of constraints. After all, considerations of cost either support options or they don't—independently of whether or not the moderate is pleased with the outcome. Ultimately, however, the moderate has to make a judgment about the relative weight of countervailing reasons—and the moderate may come to acknowledge that the weight of the pro tanto reason to promote the good is greater than he had originally been inclined to judge. Abandoning the appeal to cost would be abandoning the judgment that considerations arising from cost are powerful enough to outweigh the pro tanto reason to promote the good.

moderate and the extremist are in agreement that the constraint must not be violated. But such agreement masks the deeper disagreement between the two as to the *reason* why the given act of killing should not be performed. The extremist believes the act in question to be forbidden because, by hypothesis, it would not lead to the best outcome overall. The moderate, however, believes that this answer fails to bring out the distinct and far more important reason why the given act of killing is forbidden, a reason so compelling that it grounds a *constraint* against such acts. In principle, it should be possible for the moderate to isolate that reason even in the typical cases in which the moderate and the extremist have no dispute over whether the killing is permitted.

Nonetheless, there is a clear danger of confusion if we focus on a case in which violating the constraint will not bring about the best outcome. We run the risk of misdiagnosing the source of the moral prohibition against killing in such cases. In the attempt to isolate a reason capable of grounding a constraint against killing, we may mistakenly produce instead variants on the already-noted fact that the given act of killing will not lead to the best outcome available. (And *this* consideration, obviously, cannot justify a constraint against killing, which would prohibit such acts even when they *would* lead to the best outcome.) At the very least, by focusing instead on cases in which the extremist and the moderate disagree about the permissibility of killing, we may hope to bring the grounds for constraints more sharply into relief, better enabling the moderate to articulate and defend his belief.

Let us, therefore, imagine a case in which it *would* promote the good to kill. Imagine that by killing one innocent person I can prevent the murders of two other innocent persons. That is, by violating the constraint against killing, I prevent two other violations of the same constraint, and this is the only way to promote this desirable outcome. In short, if I kill the one, I save the two.¹⁹

Several features of this case deserve comment. Most importantly, the description is obviously schematic; I have deliberately neglected

¹⁹ Since the present discussion is meant to be neutral between a constraint against doing harm and a constraint against intending harm (see n. 14), it may be best to assume explicitly that, in the case being described, my killing the one is an essential means to (and not a mere side-effect of) my saving the two. (For a related discussion of a similar schematic example, see Scheffler, pp. 80–114, especially pp. 84–90, 98–101.)

filling in most of the details of the case. This is not because such details cannot have any moral significance. On the contrary, it is quite possible that filling in a concrete example would introduce a variety of *additional* morally relevant considerations. That is the very reason why I have left the case schematic. If the moderate is going to justify a constraint against killing, it will not help to have our attention distracted by accidental features that may muddy the moral waters. This is true regardless of whether the additional factors would happen to make the given act of killing even worse, or serve to make it somewhat more acceptable in the given case. It may, therefore, help us to focus on the constraint against killing, if we simply stipulate that in the case we are considering there are no other morally relevant factors (no mitigating side-effects, no more desirable course of action open to me, and so on).

What the case does involve, of course, is a situation in which, by hypothesis, a greater amount of good will be done by killing than by refraining from killing. This description assumes, obviously, that it is worse—all other things being equal—if two innocent people die than if only one innocent person dies. But it would be hard to deny this claim. (Matters might be different if one or more of the three people were morally culpable, or somehow responsible for being in the situation; that is the reason for imagining all three to be innocent. Of course, many other differences in the individuals might be relevant as well; but let us assume that the three are, in the relevant respects, similar.) However, since we are leaving open the possibility that a pluralist account of the good should be accepted, we cannot assume that the loss of innocent life is the only factor relevant to the goodness of the two possible outcomes.

On some accounts, for example, it might be suggested that the existence of an act of murder makes the world bad above and beyond the badness of the resulting loss of innocent life. On such a view, even if killing the one were the only way to save the two from dying of natural causes, we could not automatically assume that killing the one would result in the greater amount of good overall. For the choice would be between, on the one hand, the badness of two innocent deaths and, on the other hand, the badness of one innocent death *plus* the additional badness of an act of murder. It would be possible that the contribution that an act of murder makes to the badness of outcomes is so great that murdering the one so as to save the two from dying of natural causes would not actually promote the greater good.

I do not propose to examine the plausibility of such views. In fact, I have deliberately described our case in such a way as to sidestep the difficulty. We are not imagining that, by killing the one, I am merely saving the two from death by natural causes. Rather, I prevent the two from being murdered themselves, let us say, by two other potential murderers. Even if the existence of an act of murder makes outcomes bad in its own right, since my act of murder prevents two other acts of murder, it will remain true that by killing the one I bring about the better outcome overall. (Similar remarks, or minor variations in the case, could handle other versions of the objection.)

We are imagining, then, a situation in which it is genuinely the case that killing the one so as to save two others promotes the best available outcome. I do not mean to suggest that such cases are especially common. On the other hand, I also do not mean to concede that such cases never arise. (In point of fact, I believe that they do sometimes arise.) For our purposes it should not matter whether such cases actually occur or not. The moderate needs to defend a constraint against killing. Such a constraint would rule out killing even in cases where it would lead to the best consequences overall, so it should rule out killing the one to save the two. And this is just as the moderate would have it. He believes that it *is* morally forbidden for me to kill the one, even though this is the only way to save the two. But the moderate needs to defend this belief.

If I kill the one, I save the two. What reason can the moderate give for prohibiting me from doing this?²⁰

1. Moderates typically try to defend the constraint against killing by stressing what a horrible thing I am doing to the victim if I kill

²⁰ If the arguments criticized in this section appear to be straw men, the blame should fall upon the moderates themselves, for they have made quite a poor showing at defending their views. Most moderates are unusually vague or sloppy about the basis of constraints (even by the standards of moral philosophy); even when they are not, there is rarely much offered by way of *justification* of their views. Often they merely present their particular moderate position, at best appealing to intuition for support. Thus, e.g., Nagel in 'War' seems to describe a particular constraint without *arguing* for it (except perhaps for an implied appeal to intuition). Similarly, one looks in vain in Dworkin's *Taking Rights Seriously* or Melden's *Rights and Persons* for something like an explicit defense of constraints. There are occasional comments which seem germane, but it is hard to tell if such authors take themselves to be offering arguments or not, and if so what these might be. In this section I have attempted to construct arguments based on comments made by moderates—making the assumption (which may not always be fair) that the comments are *meant* to be/contain arguments.

him. And it is indeed a horrible thing to be killed. But it does not seem that emphasizing this point brings the moderate any closer to justifying the constraint. The death of my victim is a horrible thing to happen to him; but the deaths of the two other potential victims are obviously horrible things to happen to *them*. My victim is uncompensated—but the two others would be uncompensated too. If my victim's death is a bad thing, then the deaths of the two would, together, be worse. Why should I be forbidden to minimize the badness?

As long as the moderate sticks to what happens to a person when the constraint is violated he gets no closer to justifying the constraint. Talk of the value of an individual's being able to shape his own life, the disvalue of a life being violated, and the like—none of this helps explain why a constraint against killing should be erected; it only supports the extremist's view that it is permissible to inflict one such harm in order that a greater number of such harms—a greater overall amount of badness—may be avoided. The barrier to promoting the best outcome created by a constraint is inexplicable in terms of the badness of what happens to the victims.²¹

2. The inadequacy of stressing the badness of what *happens* to the person when the constraint is violated may suggest shifting the focus of the moderate's argument to the *relationship* which obtains between the agent and his victim.²² Moderates have often emphasized the quality or character of the acts they wish to constrain—as distinct from merely noting their outcomes. If I kill my victim, I stand in a particular relationship to him—and moderates have stressed how horrible it is for that relationship to obtain. In killing my victim, even for a good end, I am degrading him, violating his integrity as a person, treating him as a means, breaching our common humanity, and the like. Presumably I do not stand in these horrible relationships with all those whom I merely allow harm to befall.

Once again, however, the moderate's argument fails to bring him any closer to justifying the constraint. For however horrible the

²¹ Moderates who seem to offer arguments of the sort criticized here include Nozick, pp. 48–51, and perhaps 33–4; Mackie, p. 355; Fried, p. 34; and Dworkin, pp. 272–3.

²² The criticisms in the previous paragraph are also made by Nagel in 'Libertarianism', pp. 143–4, as is the suggestion that the defense of constraints must turn on the relationship between agent and victim. But he does not go on to *offer* such a defense.