

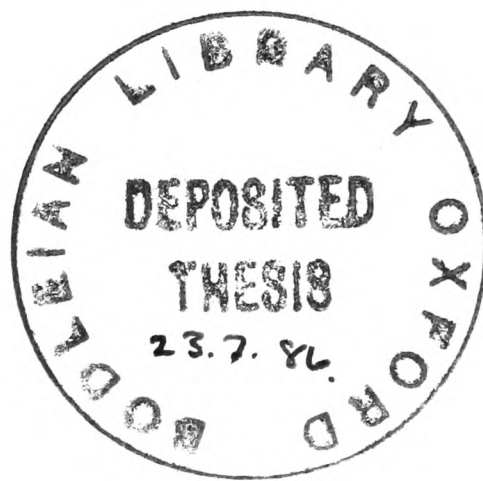
WHY SHOULD I BE MORAL? ✓

D.Phil. Thesis

by

Bradford Williamson Hooker
St. Anne's College

Hilary Term 1986⁷⁷



WHY SHOULD I BE MORAL?

Thesis submitted for the degree of D.Phil.

by

Bradford Williamson Hooker
St. Anne's College

Hilary Term 1986

ABSTRACT

I begin my discussion of the question 'Why should I be moral?' by drawing distinctions both between possible different senses of 'moral' and also between different conceptions of what morality requires. I then criticize the idea that one should be moral because it serves self-interest. Self-interest is served by one's having benevolent concern for only a fairly small number of others, but being moral involves more than this. Furthermore, having moral dispositions other than benevolence is in one's interest only if these dispositions are required by the moral code predominant in one's society. Moreover, even if we confine our attention to people who live in such a society, each person would probably be better off with moral dispositions that were not so strong that they would always get their way, but the completely moral person would presumably have overriding moral dispositions. Finally, having the correct moral beliefs may not be in one's interest. But whatever the gap between self-interest and morality, might one not have most reason to be moral? Derek Parfit has recently argued that the view that one has most reason to do whatever best achieves one's present aims (and these may sometimes be moral aims) is at least as good as the view that one has most reason to do what best promotes one's own long-term good. I attack some of his arguments. But I then go on to argue that moral requirements as such—i.e., independently of whether they are reflected in present desires—do generate reasons for action. But are these moral reasons always stronger than reasons of other kinds? On the basis of an example I describe in the closing pages, I reluctantly conclude that they are not.

ACKNOWLEDGEMENTS

Richard Hare, James Griffin, and Derek Parfit have been very generous and encouraging supervisors of my work. I owe a special debt to Parfit, because I have learned so much from his own work, and because as my main supervisor he read through many drafts without complaining. In addition, Mark Rutherford, David Mitchell, and Madison Powers gave me acute written comments on early drafts of different chapters. And I am extremely grateful to Robert Gay for his extensive comments on a complete draft of the thesis which I had, until I read his comments, planned to submit. Furthermore, I have benefited greatly from discussion with, in addition to those mentioned above, Francis Hawkings, Eric Rakowski, Michael Smith, and Roger Crisp. And I have received helpful correspondence from Gregory Kavka and Mark Overvold. Thanks are also due to Jane Hooker for various bits of proof-reading and emotional support, and to Alice and Henry Hooker for remarkable financial and emotional support.

CONTENTS

PART ONE: SETTING THE STAGE

- | | |
|-----------------|---|
| 1. Introduction | 1 |
| 2. Being Moral | 6 |

PART TWO: SELF-INTEREST AND MORALITY

- | | |
|---|----|
| 3. The Paradox of Self-interest | 26 |
| 4. Self-interest and Benevolence | 41 |
| 5. Self-interest and Other Moral Dispositions | 53 |
| 6. Self-interest and Moral Theories | 80 |

PART THREE: WHAT DO I HAVE MOST REASON TO DO?

- | | |
|--|-----|
| 7. The Present-aim Theory versus Self-interest | 101 |
| 8. Do I have most reason to be moral? | 132 |

NOTES 169

LIST OF WORKS CITED 192

PART ONE

SETTING THE STAGE

CHAPTER 1
INTRODUCTION

This thesis explores one of the oldest and yet most intriguing questions in moral philosophy—the question 'Why should I be moral?'¹ Indeed, H.A. Prichard claims that,

Anyone who, stimulated by education, has come to feel the force of the various obligations in life, at some time or other comes to feel the irksomeness of carrying them out, and to recognize the sacrifice of interest involved; and, if thoughtful, he inevitably puts to himself the question: 'Is there really a reason why I should act in the ways in which hitherto I have thought I ought to act?'²

The first point we might notice when trying to answer 'Why should I be moral?' is that if the 'should' is taken to be a moral 'should', then obviously I should be moral.³ But, equally obviously, the question is probably not asking whether morality prescribes that I be moral. The question asked is rather: granted that morally I should be moral, should I, all things considered, be moral?

Whether or not I should, all things considered, be moral would seem to depend (at least in part) on what being moral consists in. There may be different senses of the term 'morality', and there certainly are opposing conceptions of what morality requires. If I were to launch into a discussion of the question 'Why be moral?' without first setting out these different senses and different conceptions, I would have to interrupt that discussion time and again in order to explain different senses of 'morality' and different conceptions of what morality requires. In the interest of avoiding this, I set these

out in Chapter 2.

Bishop Butler famously wrote, 'when we sit down in a cool hour, we can neither justify to ourselves this or any other pursuit, till we are convinced that it will be for our happiness, or at least not contrary to it.'⁴ This suggests that a satisfactory reply to the question 'Why should I be moral?' must take the form of claiming that being moral is in one's self-interest. But is this reply correct? My assessment of it begins with an acknowledgement, in Chapter 3, that the Paradox of Self-interest applies to most of us. The Paradox of Self-interest is that one will probably have a richer life, in self-interested terms, if one cares strongly about some ends other than one's own good, so strongly even that one is sometimes led to do what one knows to be worse for oneself. Acknowledging this paradox leads to the question: are moral dispositions and desires among these strong concerns the having of which is likely to give one a more rewarding life?

This question can be divided into two: first, is it in one's interest to have benevolent concern for others? and, second, is it in one's interest to have moral motives other than benevolence,⁵ i.e., motives such as desires not to assault, rob, break one's promises, or lie? I take up the question about benevolence in Chapter 4. My conclusion here is that, with respect to most people, having or developing a limited benevolence is, on the whole, in one's interest. I remain unconvinced, however, that this is true of having a generalized benevolence.

In Chapter 5, I assess arguments which might be thought

to show that having moral motives other than benevolence is advantageous to oneself. I completely reject some of these arguments. Others can, I think, be restructured so that they have some value; and yet, having been restructured, they actually suggest a gap rather than a coincidence between having the configuration of desires the best justified morality would prescribe and having the configuration best for oneself.

In Chapter 6, I turn from moral desires to moral beliefs. I ask which normative moral beliefs—or, which conception of what morality requires—it would be in one's interest to have. I also briefly discuss the question of which metaethical view it would be in one's interest to accept.

Acknowledging a gap between self-interest and morality points to the question: on the occasions when self-interest and morality conflict, what does one have most reason to do?⁶ But some philosophers might say that I am presuming too much if I presume that the most serious challenge to morality comes from self-interest. To be sure, one possible theory is that one has most reason to do what morality requires; and this theory is opposed by the 'Self-interest Theory', which holds that one has most reason to do what is best in self-interested terms. But there is a third contender: the 'Present-aim Theory', which maintains that one has at least as much reason to do whatever will best achieve one's present aims as one has to do anything else.⁷ Now some philosophers would say that the Present-aim Theory, not the Self-interest Theory, is the best rival to the theory that one has most reason to be moral. For example, D. Parfit has recently advanced arguments intended to show that the

Present-aim Theory is superior to the Self-interest Theory. In Chapter 7, I challenge some of Parfit's arguments.

But even if his arguments can be rejected, there remains the question: does morality provide reasons for action that are stronger than self-interested reasons? I attempt to answer this question in Chapter 8. But prior to the question 'Does morality provide the strongest reasons for action?' is the question 'Does morality, as such, provide reasons for action at all?' As a start on this question, I outline some alternative accounts of how moral requirements, normative reasons, and present desires are or are not connected to one another. A couple of these accounts hold that normative reasons can only derive from present desires. That is, these accounts in effect accept the Present-aim Theory. I argue, however, that moral requirements do yield normative reasons. So if I am correct, the Present-aim Theory is false. And I suggest that moral reasons at least sometimes do have more weight than conflicting (self-interested) reasons. This yields a partial answer to the question 'Why should I be moral?'. For we can say that, with respect to some situations, I should be moral because the normative reasons for action supplied by moral requirements are stronger than any other relevant normative reasons.

But is it true that in every situation in which moral requirements are in play the normative reasons supplied by these requirements are stronger than any conflicting reasons? Concluding that it is true would, in many ways, be satisfying and reassuring. Such a conclusion would enable us to say that, in every one of these situations, one should be moral, all things

considered. But might there not be good counterexamples to the hypothesis that reasons supplied by moral requirements are always stronger than other reasons? If there are, then sometimes the correct answer to 'Why should I be moral?' is 'Taking everything into consideration, you shouldn't.' In the closing pages of this thesis I try to determine whether there are indeed any good counterexamples.

CHAPTER TWO
'BEING MORAL'

SECTION 1

An initial difficulty with the question 'Why should I be moral?' is that there may be alternative senses of 'morality'. So I begin by distinguishing between different possible senses of 'morality'. In addition, even among people who agree that they have the same sense in mind, there may be opposing conceptions of what morality in that sense requires. But I shall not argue here that any one particular conception of what morality requires is superior to its competitors. If I did do that, then the main two parts of my discussion—(1) the part about the relation between morality and self-interest and (2) the part about the relation between morality and reasons for action—would be built on a controversial foundation. What I shall do instead is to distinguish between some of the most often discussed moral theories in the latter part of this chapter and then to keep more than one of these alternatives in play in later chapters.

SECTION 2

When someone asks us 'Why should I be moral?', we may need to know whether the person who put the question is asking for reasons for doing what is moral according to the best justified morality, or for reasons for doing what is currently accepted in his or her society to be moral. This distinction between the best justified morality and the morality currently accepted by most of the people in a society might be called the distinction

between the best justified morality and the positive morality.¹ Of course, there may often be cases in which the best justified morality and the positive morality agree. In such cases the distinction between best justified morality and positive morality might seem otiose. But, though in these cases there is no difference in the extensions of these terms, there is still the difference in the terms' meanings.

Some readers might be uneasy about this distinction between best justified morality and positive morality. They might think that, if an individual disagrees with the rest of the society about what morality requires, then the individual must be wrong, on the grounds that morality is essentially rooted in a shared understanding.² But, at least in our pluralistic society, different groups disagree with one another about right and wrong. As H. Sidgwick observes, 'the code [of Public Opinion, enforced by social penalties] is manifestly fluctuating and variable, different at the same time in different classes, professions, social circles, of the same political community.'³ When different groups disagree in this way, one thing the individual must do is to try to make up his or her own mind which of the groups—if any—is correct.

But have we dismissed too quickly the doctrine that the ultimate rational grounding for morality lies in the shared understandings, implicit contracts, tacit but actual conventions? Well, many theorists as well as laymen reject this doctrine.⁴ Furthermore, even if we were to agree about morality's first principles, we might nevertheless disagree about what derivative principles follow from these shared first principles.⁵

Hence, whether or not there must be, at the root of morality, explicit or implicit agreement, we can make the distinction between what is truly morally required and what the society thinks morality requires. And because the distinction can be sustained, 'Why should I be moral?' could mean either 'Why should I do what the best justified morality actually requires of me?' or 'Why should I do what the positive morality requires of me?' In later chapters I shall be investigating answers to both these questions.

SECTION 3

Having distinguished between best justified morality and positive morality, we can move on to other ambiguities in the expression 'being moral'. We can distinguish between broad and narrow senses of 'morality' and 'moral' (or 'ethics' and 'ethical').⁶ Consider the idea that moral principles are to be defined as whatever principles are supremely important, or overriding. On this definition of morality, whatever one has most reason to do, all things considered, is what morality prescribes. Let us call this: the broad sense of the term 'morality'. (Some people may be sceptical that there is the broad sense of the term 'morality'. For my purposes, as I shall later explain, I do not need to insist that they are mistaken.)

The narrow sense of 'morality' is more difficult to define than the broad sense.⁷ Consequently, many writers start off by saying what narrow morality is not, or what it contrasts with: e.g., prudence, etiquette, aesthetics, and perhaps law and codes of honour.⁸ But what distinguishes the

considerations of narrow morality from considerations of these other kinds? One possible answer is that a principle of narrow morality must meet certain formal restrictions other than or beyond the single formal restriction met by principles of broad morality, viz., overridingness.⁹ An alternative view is that the principles of narrow morality are distinguished by restrictions on their content.¹⁰ For example, it might be thought that moral principles are those that would be chosen by someone with impartial concern for all. Or it might be said that narrow morality has a particular purpose or role, e.g., to promote social harmony (though narrow morality is not the only institution to have this role).¹¹ A more specific view along the same lines is that the principles of narrow morality have the function of helping us overcome pure co-ordination and, particularly, partial-conflict problems.¹² Alternatively, it might be held that there is nothing more general to be said than that the principles of narrow morality have to do with particular kinds of behaviour--for example, assault, stealing, promise-breaking, lying, unfairness, coercion, disloyalty, and behaviour which manifests a failure to give special consideration to the welfare of people who stand in special relations to us. Someone who maintained this view of narrow morality might hold that it is a mistake to think that we need one principle, such as a principle about impartial concern or the promotion of social harmony, to underpin or justify the various different principles and norms of narrow morality. For example, on this view, the principle that one ought to give greater weight to the welfare of one's own family and friends would be considered to be

itself basic, i.e., not derived from some putatively more fundamental principle, such as a principle about impartial concern.¹³

It would be nice to have an abstract statement of the distinguishing mark of narrow morality that advocates of these different views could all accept. But, I know of no statement of the concept of narrow morality which embraces all the systems we can clearly recognize as narrow moralities. Nevertheless, we sometimes need not be able to define a thing in order to tell that it is different from something else. And clearly there is a difference between narrow morality and prudence, aesthetics, etiquette, and so on. Moreover, no matter which conception of narrow morality of those adumbrated above we choose, narrow morality is different from broad morality.

The distinction between broad and narrow morality and the distinction between best justified and positive morality are not to be conflated.¹⁴ In fact the two distinctions cut across one another. That is, we can distinguish: (i) best justified broad morality, (ii) best justified narrow morality, (iii) positive broad morality, and (iv) positive narrow morality. Now, in many cases, one of these four concepts will have the same extension as that of one or more of the other concepts. For an example of a case where two of these concepts have the same extension, imagine a society in which the principles of the positive narrow morality are thought by the members of the society to provide the strongest reasons for action. Here, the principles of the positive narrow morality are also the principles of the positive broad morality. And it might be the

case that—indeed it is a profound question whether—the extensions of best justified narrow morality and best justified broad morality are the same, i.e. that what we are morally required to do (in the narrow sense of 'morality') is what we have most reason to do, all things considered. I shall take up that question in my final chapter. But the important point here is that the four concepts distinguished above are different.

And whether or not the extensions really do differ, we can describe a case in which someone believes that they do. Consider a society in which virtually everyone privately believes, but does not publicly profess, that at least sometimes they have better, stronger reasons for disobeying the positive narrow morality than for obeying it. In such a society there would be a difference between the principles of the positive broad morality and the principles of the positive narrow morality. Now suppose that I live in this society, and that what I believe morality requires me to do, in the narrow sense of 'morality', is not what I believe I have most reason, all things considered, to do. Then, as we might put it, my narrow moral beliefs do not coincide with my broad moral beliefs. Suppose furthermore that my narrow moral beliefs conflict with the narrow moral beliefs current in this society, and that my broad moral beliefs conflict with the broad moral beliefs current in the society. I might then think that the extensions of the four concepts are different.

SECTION 4

Having distinguished from one another best justified broad morality, best justified narrow morality, positive broad morality, and positive narrow morality, we can now ask which of these four concepts is important to the question why one should be moral.

When people ask 'Why should I do what I think morality requires?' they are usually asking what reasons (if any) they have for doing what the best justified narrow morality demands. They are tempted to act on considerations--often self-interested ones--that conflict with the action-guiding considerations the best justified narrow morality furnishes, and they want to know why they should give priority to the latter. They want to know why, or rather if, the considerations this morality generates are stronger than other kinds of action-guiding considerations. This is an intelligible and not uncommon question. It is the question Sidgwick faces at the end of his Methods of Ethics (without finding a wholly satisfactory answer), and it is a question to which a considerable portion of this thesis will be devoted.

But, we would have difficulty making sense of 'Why be moral?' if we were to interpret it as a request for reasons for conforming with best justified broad morality. There is nothing more to acting in accordance with best justified broad morality than acting on whichever prescriptive system generates the strongest reasons for action. Thus asking what reasons one has for doing what one believes best justified broad morality prescribes manifests a lack of understanding of the concept of

'broad morality'. In reply to this, the questioner might ask, 'Well, why should I do what (I believe) I have most reason to do?' Here we can only reply: 'in pointing out what you have most reason to do, we have exhausted your strongest reasons for action. So you cannot now go on to ask us to supply further reasons for action, though of course you can decide to be irrational, that is, to ignore considerations that you yourself believe constitute reasons for action.¹⁵ Another way of putting this point is to say that while asking what one has most reason to do is a sensible question, asking why one should do what one has most reason to do is not.

So, to read 'moral' as referring, in the question 'Why should I be moral?', to the best justified broad morality makes the question into 'Why should one do what the best justified broad morality requires?'. As we have just seen, this is not a sensible question. I shall therefore not be interpreting the term 'moral' as referring to best justified broad morality. That said, there should be no confusion if, in order to save words, I henceforth use the shorter expression 'best justified morality' to mean 'best justified narrow morality'.

Let us turn now to the two sorts of positive morality. Ultimately, the reason positive morality is relevant to the question 'Why should I be moral?' is external sanctions. By 'external sanctions' I mean legal and social sanctions. Sanctions can be negative or positive. Negative legal sanctions are fines and punishments. Negative social sanctions are such things as resentment, distrust, ostracism. Positive sanctions are legal or social rewards. Now all these sanctions generate

self-interested considerations in favour of conforming (or at least seeming to conform) with whatever evaluative principles are backed up by external sanctions. But in a society where the principles of the positive broad morality are not the same as the principles of the positive narrow morality, which principles are backed up by sanctions? My inclination is to say that the sanctions would be attached to only the positive narrow morality. But rather than try to defend this claim here, I shall leave open the question of whether principles of the positive broad morality might also be backed up by social sanctions. So the distinction between positive narrow and positive broad morality is not what is important; rather it is the distinction between the society's evaluative principles that are backed up by sanctions and the society's principles that aren't. I shall accommodate this point by henceforth using 'positive morality' to mean whatever norms and principles are backed up by social sanctions. All of positive morality, defined in this way, is relevant to the question 'Why, from the self-interested point of view, should I be moral?'

SECTION 5

Having said which senses of 'morality' I shall be considering, I turn now to distinctions between different substantive moral theories, that is, theories about what morality requires. One of the most fundamental distinctions between such theories is that between consequentialist and deontological theories.

Act-consequentialist theories insist that an act is morally right if no alternative act would produce more value, or better

consequences. Act-utilitarianism, of course, maintains that what has value is welfare and that the right act is the act that maximizes welfare. Now if act-consequentialist theories are defined as theories holding that the morally right act is that which produces the greatest amount of non-moral value,¹⁶ act-utilitarianism, with its concentration on welfare, might well appear to be the only possible act-consequentialist theory. But 'consequentialism' has been used by many recent writers in a way that is capacious enough to include theories in which consequences are to be evaluated by how much moral as well as non-moral value they contain.¹⁷ What might these moral values be? One good candidate for inclusion in the consequentialist maximand is equality. And even beyond that, consequentialism need not restrict itself to values that would come about after the act, as features of the outcomes of acts. That is, there is room for consequentialists to hold that some acts have not only extrinsic but also intrinsic moral value, that how well the world goes is partly constituted by what acts occur. Thus some consequentialists might evaluate how the world goes not only, positively, by how much welfare and equality it contains, but also, negatively, by how many morally bad acts of (e.g.) murder, torture, assault, robbery, blackmail, promise-breaking, and lying occur.¹⁸ These consequentialists might affirm that one possible history of the world might be worse than another simply because it contains more acts of torture than does the other, even if it also contains greater welfare and equality. To be sure, many other consequentialists would deny that torture has intrinsic disvalue. But, however

much different consequentialists disagree amongst themselves about what has intrinsic value and what has intrinsic disvalue, all act-consequentialists agree that an act is right if there is no other available act that would cause a better history of the world.

Deontologists disagree. They insist that one is forbidden to do some acts even when doing them would produce the best history of the world.¹⁹ A consequentialist who thought that one of the things with intrinsic disvalue was murder might say that, other things being equal, we ought to minimize the total number of murders, even when we must ourselves commit murder(s) to do this. Deontologists, by contrast, claim that it is wrong to commit murder even in those cases where one's committing murder would result in there being less total murders. The same sort of claim, mutatis mutandis, would be made by deontologists about certain other kinds of acts.

Another way of expressing the nature of this dispute is to say that it is about whether morality is correctly conceived of as containing agent-relative first principles.²⁰ Agent-relative principles are defined as those principles that include an essential (unique and ineliminable) reference to the agent in the statement of the conditions under which the agent is morally required to promote some event or outcome.²¹ We might, borrowing from T. Nagel,²² put moral principles in the following form: 'For all agents A and events or states of affairs E, if _____, then A is required to promote E'. (When E is an act, the performance of E is to count as a case of promoting E.) The conditional's antecedent of course states the

conditions under which the agent A is required to promote the event E. Obviously, all moral principles about what one should do mention the agent in the consequent of the conditional.²³ But the essential reference that distinguishes agent-relative from agent-neutral principles occurs in the conditional's antecedent. For example, the agent-relative principle 'For all A and E, if there exists a promise such that it was made by A and E is an event which fulfils that promise, then A is required to promote E' does involve an essential reference, in the antecedent, to the agent.

Why must the agent be mentioned in the antecedent of the conditional? One sort of agent-relative principle, the sort of which the above principle about promise-keeping is an instance, focuses on the relation between the agent and his or her own acts of some kind. There needs to be a reference to the agent in the antecedent to pick out these acts.

It is worth underlining that these agent-relative principles do not tell one to focus on the relation between oneself and all acts of the kind in question. For example, one is told to promote the fulfilment of one's own, not of everyone's, promises. In contrast, act-consequentialists who hold that promise-keeping has intrinsic value would say that it can be morally right to break one's own promise in order to cause promises made by other people to be fulfilled.

A second sort of agent-relative principle focuses on the relation between the agent and particular people to whom the agent stands in special relations, e.g., family, friends, benefactors, colleagues, fellow-citizens, etc. Since such

principles focus on the relation between the agent and certain people related in some way to the agent, they must mention the agent in order to refer to these people. Consider the following principle as an example: 'If E would benefit a member of A's own family somewhat less than some other event would benefit someone else, then A is required to promote E rather than the other event.' The idea here is that, when deciding what to do, the agent is morally required to give greater weight to the welfare of his or her own family members (or friends, benefactors, clients, etc.) than to the welfare of others. Contrast an agent-neutral principle about family ties: 'For all A and E, if there exists a person such that E is an event in which that person gives somewhat greater weight to the interests of his or her own family members, then A is morally required to promote E.' Such a principle contains no essential reference to the agent in the antecedent. (Of course there may be a situation where A could give special weight to the welfare of his or her own family members, in which case the principle requires A to do this.)

The disagreement I have been discussing between deontologists and consequentialists is over first principles. As I explain in the following section, consequentialism might well tell us to have, at some instrumental level, dispositions to obey agent-relative principles, because our having these dispositions will produce better consequences. It may turn out that there is not a great divergence between the agent-relative principles that the most plausible deontological views defend and the instrumental agent-relative dispositions

that are endorsed by the most plausible versions of consequentialism. But, however that may turn out, there is still a disagreement between deontologists and consequentialists over first principles.

There may be many possible deontological views, but I shall concentrate on one, Common-sense Morality. This morality gives us 'negative duties' not to harm any undeserving person even when by doing so we could produce better overall results. (It may also forbid us to do certain things—e.g., tell lies, break promises—even when doing them would harm no one.) It also gives us positive duties to aid. These positive duties are to help anyone when the cost to oneself of doing so small in comparison to the benefit conferred on the other person,²⁴ and to give special weight to the interests of those to whom we are connected in certain ways—family, friends, benefactors, colleagues, etc.²⁵

SECTION 6

I want now to mention some complications within consequentialism. One complication is its multi-level structure. Some consequentialists distinguish between the fundamental goal or goals of morality and the instrumental desires, dispositions, and rules the having of which will best achieve the fundamental goal or goals.²⁶ For example, a consequentialist who believes that the ultimate moral goal is the maximization of net happiness might hold that if our predominant desire were to maximize happiness, the world would not contain as much happiness as it would under certain other conditions.

One way in which better consequences might be produced by our not having the disposition to do always what will produce the best possible consequences is as follows. If we had no inhibitions about doing certain kinds of act, we would often, as it turns out, not do what will produce the best consequences, impartially considered. Suppose that I, like many other people, have great difficulty in being truly impartial when I am calculating the expected benefits of alternative lines of action. For example, I systematically (though not intentionally) underestimate the expected harms to others of possible outcomes that would benefit me, and overestimate the expected harms to others of outcomes that would harm me. Now let us suppose that I could develop in myself either of two alternative dispositions. I could become disposed to steal only when my stealing something would produce what I believe to be very much better consequences, impartially considered. Or I could become disposed to steal whenever I believe this would produce better—even if only slightly better—consequences, impartially considered. Now the consequentialist might say that, because my calculations about consequences tend to be biased, the acts I would perform if I had the first disposition would probably have worse consequences than the acts I would perform if I had the second disposition. Note that the first disposition would be better than the second, according to consequentialism, because it would lead me to perform acts with better consequences than the acts the second disposition would lead me to perform.

Another complication with consequentialism arises from the following considerations.²⁷ Suppose I ask myself what

projects I should take up in order to bring about the best consequences, given that most other people are in fact quite selfish. I reason that I should devote myself exclusively to improving the plight of the world's worst-off. Now I ask myself what concerns I should cultivate in myself. I reason that I should not cultivate any special attachments (that is, start a family or form friendships) because having such attachments would get in the way of my devoting myself exclusively to improving the plight of the world's worst-off (unless, of course, the attachments were to the worst-off).

But my not having such attachments might very well make my life go worse for me than it would go if I did have such attachments. So I might object, 'Why should I have the set of desires and concerns that, given the way others are in fact disposed, will bring about the best consequences impartially considered, when my having this set will be bad (in prudential terms) for me?' I am being told to be extra altruistic and self-sacrificing in order to make up for the selfishness of others—but this is unfair to me.

Act-consequentialists hold that it is wrong not to cause oneself to have the set of desires and concerns the having of which will produce the best consequences, given the predominant selfishness of most others. But the thoughts I rehearsed in the last paragraph give rise to a form of consequentialism which tells each of us to have one of the sets of desires and concerns which are such that, if everyone had one or another of these sets, the best consequences would be produced. Let us follow Parfit in calling this theory Collective Consequentialism.²⁸

But what are these sets of desires and concerns which are such that, if everyone had one or another of them, the best consequences would be produced? First of all, let us admit that virtually everyone cares more about some people than about others. In other words, virtually everyone falls short of complete impartiality. Even beyond this, it is presumably true that being both equally and deeply concerned for everyone alive today (much less for everyone who will ever live) is, for at least most people, a psychological impossibility.²⁹ It therefore appears that if people are to continue to have the deep attachments they now have, they cannot become completely impartial; and if they were to become impartial, they would completely lose their deep attachments.

This is not to deny that people could perhaps become more impartial than they are now without losing their deep attachments. So, for the sake of argument, let us contrast (i) a possible future in which people are that much more impartial with (ii) a possible future in which people are completely impartial. Given the psychological limitations mentioned, in the first alternative future each person would have strong benevolent concern for a few people plus fairly weak benevolent concern for the rest; in the second, each would have fairly weak benevolent feelings towards everyone.

Now let us add a further crucial assumption: much of people's happiness comes from their deep attachments, and without these deep attachments there would be much less happiness. So a future in which people had some deep attachments would (even at the cost of their not having equal concern for everyone) contain

much more happiness than would a world in which absolutely no partiality (and so no strong attachments) existed. If this is correct, and if there is no other benefit which could compensate for the loss in happiness, then the better of the two alternative possible futures would be the one in which each person had strong concern for some others at the cost of not having equal concern for everyone. Given the way Collective Consequentialism tells us to think about our patterns of concern, Collective Consequentialism would require each of us to be like the people in this better future.

There are of course serious objections to Collective Consequentialism, perhaps the most serious of which is that on the one hand it is consequentialist and on the other, because it allows us ignore what others are actually doing, it can lead us to do what will in fact have disastrous consequences. Act-consequentialism is of course also open to serious objections, one of which is the objection mentioned above that it unfairly asks one person to compensate for the selfishness of others. It is not my task here, however, to try to determine which is the best moral theory. The task is rather to outline the main alternative theories in order to set the stage for later discussion.

SECTION 7

Another distinction that needs to be mentioned is the distinction between reciprocal and unilateral conceptions of what morality requires.³⁰ Roughly, reciprocal conceptions hold that we are morally obligated to treat others well only if they are treating,

or will treat, us well; unilateral conceptions of morality require us to treat others well even if they do not reciprocate.

The distinction between reciprocal and unilateral conceptions may be related to the distinction within our positive morality between the ordinary requirements of morality and the principles of supererogation.³¹ That is, we might well think that our positive morality does not require us (e.g.) to go out of our way to help those who refuse to help us, or to keep our promises to people who break their promises to us. Such acts would be considered, not obligatory, but supererogatory. Still, whether or not that claim is correct, clearly our answer to the question of whether being moral is in a particular person's interests may turn on whether we have in mind a unilateral or reciprocal conception of morality. For there might be prudential reasons for conforming with one of these conceptions of morality which are not reasons for conforming with the other.

I should add that, in the rest of this thesis, though from time to time I shall bring into the discussion unilateral conceptions of morality, I shall for the most part ignore supererogation as such. 'Why should one be moral?' is a difficult enough question if being moral is taken to consist in obeying merely the ordinary requirements of morality. I shall thus make no attempt to determine whether one should be moral in the sense of doing what is morally supererogatory.

SECTION 8

To summarize: in this chapter I have considered various things that might be meant by 'be moral' in the question 'Why should I be moral?' After distinguishing between a broad and a narrow sense of 'morality', and between best justified and positive morality, I argued that we should not interpret the question as 'Why should I conform with best justified broad morality?'

Instead, the appropriate interpretations are, 'Why should I do what the best justified narrow morality requires of me?' and 'Why should I do what the positive morality requires of me?' I then outlined some competing normative moral theories. All this helps set the stage for the discussion (in the next part of the thesis) of the relation between self-interest and morality, and for the discussion (in the final part) of the relation between morality and reasons for action.

PART TWO

SELF-INTEREST AND MORALITY

CHAPTER 3

THE PARADOX OF SELF-INTEREST

SECTION 1

Perhaps the reason one should be moral is that being moral is in one's self-interest. In this Part of the thesis I consider this proposal. In the present chapter I make some preliminary points about self-interest.

I shall follow the practice of taking 'prudence' and 'self-interest' to be equivalent.¹ So to act prudently, or in one's self-interest, is to do what will make one's life go best in self-interested terms, in other words, to do what will maximize one's own good, or welfare. By 'one's life', I mean one's whole life, from start to end. Thus it is imprudent to choose a lesser present benefit to oneself over a greater future benefit (or, indeed, a lesser future benefit over a greater present benefit). Of course, future goods are often less certain and may be briefer. Naturally one should take into account certainty and duration when estimating the benefits and losses one might receive from alternative courses of action. So, being precise, we should say that it is imprudent to choose a course of action with a smaller expected net benefit to oneself over one with a greater expected net benefit, even if the smaller expected net benefit will come sooner than the greater expected net benefit. Having acknowledged this point, I shall from now on often save words by using the terms 'benefit' and 'loss' to stand for 'expected benefit' and 'expected loss'. Let me add that my discussion will be in terms of benefits and losses with

natural causes—I set aside the possibility that one might be rewarded or punished, during life or after death, by some supernatural being(s).

SECTION 2

Consider the doctrine of Psychological Egoism. Psychological Egoism, in any form worth considering, makes an empirical claim about our actions and desires. It claims that every intentional human action is guided by the desire that one's life go as well for oneself as possible. (It will be helpful to have a label for this desire: 'self-love' will serve this purpose.)

There may be different versions of Psychological Egoism. One version might state that each of us has at bottom only one desire, that his or her life go as well in self-interested terms as possible. Call this the 'Single-desire Version'. Much of our experience contradicts this theory. To be sure, experience may suggest that most people care strongly about their own welfare. But, of course, experience also indicates that people have many other desires—desires for food, sex, affection, achievement, self-respect, to see beautiful things, to know the truth, and so on. This multiplicity manifests itself in the familiar phenomenon of conflict between one's desires. There could be no intrapersonal conflict if one had only a single desire. Moreover, the instance of conflict that is most telling in the present context is the conflict between self-love and one's other desires. There would be no room for this conflict if self-love were one's only desire.

It would be a mistake to reply on behalf of the

Single-desire Version that, although each person seems to have many different desires and they sometimes seem to conflict, all we really care about is our own pleasure, and that the fulfilment of our various other desires is the means to that. True, we often derive pleasure from obtaining the objects of these other desires. But these desires are for their own objects, not for the pleasure that may be a by-product of getting those objects. Again true, most of us also care very much about our own pleasure, and this concern often prevails over our desires for other things when we think that getting those other things would have unpleasant consequences. Still, pleasure is a concomitant of the objects of the other desires, not the object itself.² So it is false that all we really care about is our own pleasure.

A second version of Psychological Egoism accepts that we do have other desires but maintains that our strongest desire is always the desire that things go as well as possible for ourselves. Call this the 'Strongest-desire Version'. This version does not make the absurd mistake of supposing that we do not desire food, sex, affection, achievement, and the rest. But it does suppose that, when these desires come into conflict with one's long-term good, the desire for one's long-term good (i.e. self-love) always prevails. This idea, however, is also incorrect. As Bishop Butler laments, 'nothing is more common, than to see men give themselves up to a passion or an affection to their known prejudice and ruin, and in direct contradiction to manifest and real interest, and the loudest calls of self-love'.³ In conclusion, then, Psychological Egoism, either in its Single-desire Version or in its Strongest-desire

Version, is false.

Let me explain here why I have bothered to mention that the Strongest Desire version of Psychological Egoism is false. I shall soon be entertaining the hypothesis that it is in one's best interest not to care most about what is in one's best interest. But if the Strongest Desire version of Psychological Egoism were true, our psychology would be such that we would be unable to care less about self-interest than about other things. If we were unable to do this, it would be pointless to entertain the hypothesis mentioned. So, in order for the hypothesis to be worth discussing at all, it must be established that caring less about self-interest than something else is indeed psychologically possible.

SECTION 3

In his Enquiry Concerning the Principles of Morals Hume suggests that acting with benevolence and kindness is naturally pleasing to us, that it 'keeps us in humour with ourselves as well as others'.⁴ This might be taken to imply that the demands of morality and self-interest coincide, as Hume officially holds.⁵ But even if we assume that there is 'some particle of the dove kneaded into our frame',⁶ the question of whether self-interest tells us to be moral is still open. Assuming our desires and dispositions are to some extent plastic, should we retain our benevolence and other moral dispositions or should we try to extract the dove from our frame?

But are our desires plastic enough for this to be a real question? One cannot just decide to have, at this moment,

different desires. Rawls observes, 'We act now as the sort of person we are and from the wants we have now, and not as the sort of person we might have been or from desires we would have had if earlier we had only chosen differently.'⁷ Nevertheless, acknowledging that one cannot choose what present desires to have does not impede us from observing that, in Rawls' words, 'we can certainly decide now to do something that we know will affect the desires that we shall have in the future.'⁸ To take a particularly clear example, some young people may be well aware that their choice of school and career will determine what desires and aspirations they will come to have later in life.⁹ (As I have already suggested in Chapter 2, there may be limits to the kinds of self-transformation that are empirically possible.¹⁰ This is something to keep in mind when discussing the question of what pattern of concern it would be best in self-interested terms for an agent to try to have.)

So since we can take steps to change the desires we have, there is the question of what desires it is prudent to have or develop. Someone might ask whether it wouldn't be prudent to rid oneself of all one's desires except self-love. The answer to this question is 'no'. Reflection on the plausible theories of self-interest shows that one needs these other desires in order to prosper.

Consider first the Desire-fulfilment account of self-interest. This is the theory that the one's life goes well just to the extent that one's desires are fulfilled. This theory takes into account the number, duration, and intensity of all the desires or preferences a person has, or might have had, at one

time or another in life. On the Desire-fulfilment Theory, the fulfilment of one of my desires counts as benefiting me even if I do not know that it has been fulfilled, and even if the fulfilment of this desire consequently does not bring me any happiness.¹¹ (Different forms of Desire-fulfilment Theory differ about whether 'global' desires about the whole of one's life, e.g. the desire that one live an honest life, should be counted more heavily than 'local' desires about the immediate situation, e.g. the desire that one have some drug now.¹² But what I shall be saying in this section will be true no matter which form of the theory is best.) Now suppose I held the Desire-fulfilment Theory of welfare and simultaneously had only one desire: self-love, or the desire that I have the best possible life, in self-interested terms. But, as things stand with me, this desire cannot be acted upon. A desire for the maximum fulfilment of my desires implicitly refers to desires other than itself.¹³ So, acting on this desire necessitates my having other desires as well. (A desire about one's other desires, such as the desire that one's other desires be fulfilled, can be called a 'higher-order desire'.¹⁴)

We should grant, however, that the ability to act on this desire does not necessitate that I have these other desires at the present moment. Suppose I know now that if I come upon a dramatic sunset in the future, I shall find myself liking the experience, though I will not have had until that moment any desire to see dramatic sunsets. I could now act so that in the future I shall come upon dramatic sunsets. In this case, I would be acting now on my desire for the maximum fulfilment of my

desires, and yet I would not have any other desire at present. But, in this example, I could not act on my desire for what is best for myself in desire-fulfilment terms if it were not true that I have, at some time, some other desire. (Indeed, some advocates of the Desire-fulfilment Theory might say that their theory implies that one ought to cause oneself to acquire however many new desires would then be maximally fulfilled.)

Now let us ask whether I could act on my desire for the best life for myself if I accept a Hedonistic account of self-interest. The most defensible Hedonistic theory of self-interest is Preference-hedonism.¹⁵ This theory maintains that the best life in self-interested terms is that in which one has more mental states that one prefers (at the times of the mental states) than one would have had if one's life had gone in any other way. (I wrote 'more mental states that one prefers', but this theory of course takes into consideration not only the number but also the intensity and duration of these desired mental states.) Now suppose I accept Preference-hedonism. Then, desiring my own greatest good would amount to desiring that I get the mental states I prefer. But I can't get the mental states I prefer unless I have preferences between mental states. Therefore, Hedonism, like the Desire-fulfilment Theory, tells me to maintain or develop desires other than the desire for my own greatest good.¹⁶

But is there some other theory of self-interest according to which one can act on one's desire for what is best for oneself without its being true that one has, at some time,

other desires? Consider the Objective-list Theory. On this theory, the best life in self-interested terms is not necessarily the one with the most pleasant mental states, nor the one involving the greatest desire-fulfilment, but rather the one richest in terms of the things on the list.¹⁷ (Examples of what might be on this list are achievement, knowledge, friendship, autonomy, the appreciation of true beauty, and pleasure.) Now, given this theory, I can act on the desire for what is best for myself without ever having any other desires. For example, an Objective List Theory might say that my acquiring (e.g.) knowledge would make my life go better even if I did not desire knowledge. So, I would be able to act on my desire for what is best for myself by seeking knowledge, even if I had no independent desire for knowledge (or anything else), whereas this is just what is unintelligible according to both Hedonistic and Desire-fulfilment Theories. On the other hand, no doubt an Objective List Theorist would agree that I am better off in self-interested terms having other desires. Pleasure is one of the things on any plausible Objective List. And, as we saw in the preceding paragraph, pleasure presupposes preferences between mental states. Thus, according to an Objective List Theorist, I ought to have desires, or preferences, other than the desire for my own greatest good, because this is a necessary condition of my getting pleasure, which is on the list.¹⁸

We might think that the Objective List Theory as stated is implausible: how can it be good for me, say, to see true beauty if I do not desire to see it and get no pleasure from it? This suggests that we either dismiss the Objective List Theory or

reform it. Reform might yield a Objective List Theory according to which what is good for someone is to have the things on the list and simultaneously to want to have just those things.¹⁹

If we accept such an improved Objective List Theory, we will again have to say that one should have desires other than the desire for the best possible life for oneself.

SECTION 4

Having established that one will be better off if one has some desires other than self-love (the desire for one's own greatest good) we face the further question of whether one is better off if these other desires are strong enough to overpower (at least occasionally) self-love. It might seem that the answer to this question is obvious. Wouldn't I probably have the better life in prudential terms if my other desires were so much weaker than my desire for what is best for myself on the whole that they *were* never able to impel me to act imprudently?

Perhaps not. Consider the so-called Paradox of Hedonism. This is the paradox that one's life is likely to be happier if one's other desires are not so much weaker than self-love that they can never get their way when they come into conflict with it. This paradox is very widely acknowledged.²⁰ Furthermore, because all plausible theories of self-interest give considerable weight to pleasure, they all would accept that one is likely to have not only a happier but also a better life if one's other desires are not so much weaker than self-love that they never get their way against it. We

could call this the Paradox of Self-interest.

It is worth taking some care over how exactly these paradoxes arise. In Butler's view, 'a person may have so steady and fixed an eye upon his own interest, whatever he places it in, as may hinder him from attending to many gratifications within his reach, which others have their minds free and open to.'²¹ This suggests that preoccupation with self-interest gets in the way of one's own enjoyment. Suppose my friends insist I join their card game. If my concentration on the game is distracted because I am frequently asking myself whether continuing with the game will afford me more pleasure than some alternative endeavor would, I will not enjoy the game as much as I would if I could devote to it my unencumbered attention. Indeed, we might find most of our activities less pleasant if we approached them with a mind divided between the activity itself and prudential calculations about whether we should be doing something else.²²

But divided attention might not be the only problem. Consider the following example. Catherine has just come up to university from her family's farm in the country's remotest region. Although she must choose a course of study and fulfil the requirements of that course of study, she is well aware that she has been delivered out of an intellectually bare area and into an environment abounding in lectures on psychology, history, literature, art, music, etc. Lucky enough to have strong desires to learn about these subjects as well as about her own chosen subject 'Metallurgy and Science of Materials', Catherine relishes spending her free time going to lectures on the many subjects

that interest her. If she did not have very strong desires to learn about these subjects, she would find her free time at university boring. Thus she gets more pleasure out of life because she has these strong desires.²³

Her life is richer also on non-hedonistic theories self-interest. Richer on the Desire-fulfilment Theory because if she did not have these strong desires to learn about the other subjects, she would have no strong interests to which she could devote her free time. Richer, too, on the Objective List Theory because she would get less pleasure, which is one of the items on the list, out of her time at university if she didn't have these strong desires.

But Catherine's strong desires sometimes conflict with, and even prevail over, her desire to do what will be best for herself. On some occasions Catherine would be better off on the whole if she could ignore her thirst for knowledge and so skip the lectures that are not relevant to her course. For example, she goes to so many extra lectures that she sometimes is made over-tired and ill. We can note that the enjoyment she gets from going to the last few extra lectures does not compensate her, in hedonistic terms, for the unpleasant effects of being ill. We can note also that the fulfilment of her desires to go to these last few extra lectures does not make up, in desire-fulfilment terms, for the lost opportunities to fulfil other desires. Thus, the pattern of desire giving her the best possible life, as understood by either Hedonistic or Desire-fulfilment Theories, will sometimes lead her to perform actions that, looked at singly, have worse outcomes for her than some alternative actions

would.²⁴

I should make explicit a crucial assumption behind my story about Catherine. I presupposed that if Catherine's various desires are strong enough to make her life rewarding, her self-love cannot be so much stronger that it always gets its way. And I shall assume that most of us are like Catherine in this respect. That is, most of us can have either patterns of concern in which self-love is so much stronger than other desires that it always gets its way, with the drawback that the other desires are insipidly weak, or patterns in which various other desires are strong enough to make life rich in hedonistic and desire-fulfilment terms, with the drawback that self-love is not strong enough always to get its way.

But what if someone objected that Catherine would be even better off if these desires enriching her life were 'fine-tuned' so that they switched off whenever they conflicted with her doing what was best on the whole for herself? This objection bids me to make explicit another assumption: as a matter of empirical fact, at least most of us could not fine-tune all of our desires (other than self-love) so that they cease to exert their pull whenever acting on them would not be in our self-interest. In Mackie's words, 'dispositions cannot be switched on and off in deference to the calculation of likely consequences on particular occasions The practical choice will be between one fairly persistent disposition and others, equally persistent, that contrast with it.'²⁵

SECTION 5

So, developing desires to such a pitch that they can overpower one's self-love can be in one's own interests. But some might think this conclusion somehow twists the truth. They might say that if, in order to make your life go better on the whole in self-interested terms, you caused yourself to have (e.g.) benevolent desires that certain people flourish, then these desires are really nothing more than the desire that your own good be maximized.

That is a mistake. Suppose that you find yourself imprisoned for life in a cell with two people you do not care about. But you think that it would make your life much more pleasant if you actually cared about them. With this in mind, you beg the prison psychologists to use behaviour-modification, hypnotherapy, or some other technique to induce in you an altruistic concern for them. Now notice that your motive for acquiring the new desire is to make your own life go better for you, but the object of the new desire is that their lives go as well as possible for them. This example clearly illustrates that the object of a desire should not be confused with the motive for acquiring it.

To drive home the point, suppose that developing this concern for the well-being of your cellmates does indeed make your life during the following twenty years much more agreeable. But it then transpires that you have an opportunity to escape. Yet this escape would involve your doing something that greatly harms your cellmates. Now we would expect your concern for their welfare to continue to exert its pull on you even after you

discover that doing what is best for them sharply conflicts with your self-interest. But the fact that the two desires now pull in different directions demonstrates that they do have different objects—the object of one desire being your own welfare and the object of the other being their welfare. It is not the case that the two desires really have the same object.²⁶

Someone might object to this by claiming that any desire one cultivates in oneself for some further reason is an extrinsic desire, whereas a truly benevolent concern for someone must involve an intrinsic concern for that person's welfare. Part of this claim is correct—a truly benevolent concern for someone does involve an intrinsic concern for that person's welfare. The other part, however, is incorrect—it is a mistake to think that any desire one cultivates in oneself for some further reason is an extrinsic desire. Briefly put, the distinction between extrinsic and intrinsic desires is as follows. If I want X as a means to something else, my desire for X is extrinsic; if I want X for its own sake, my desire is intrinsic. One way of getting at this distinction is in terms of whether a desire has a life of its own. To use a common example, my desire for a painful operation is extrinsic since I desire the operation only because I believe it will ultimately fulfil my further desire that I be healthy. And if I learned that the operation would not help my health, I certainly would no longer want it. So my desire that I have the operation has no life of its own.

Contrast this with the case of an intrinsic desire. In deciding whether a desire is intrinsic, what matters is whether the desire has now a life of its own, not whether it was

initially fostered by some other desire. Your benevolent concern for your cellmates' welfare does not extinguish the moment you find out that what would be best for them conflicts with what would be best for you. Once this benevolent concern has been cultivated, it has a life of its own. This shows that it is intrinsic. (I am not forgetting, of course, that we can take steps to rid ourselves of desires we previously cultivated in ourselves. A desire can have a life of its own without being imperishable.)

A caution: the purpose of this section was not to argue that developing benevolence is in one's interest. Rather, the point was to explain that if it is, then the benevolence is truly benevolent and not in some way surreptitiously selfish. So far, I have merely described somewhat abnormal circumstances in which (by hypothesis) having benevolent concern was, for a period of time, prudent. It remains to be seen whether, in the more ordinary circumstances, having benevolent concern for others is generally in one's own interest.

CHAPTER 4

SELF-INTEREST AND BENEVOLENCE

SECTION 1

In the previous chapter I tried to show how one's life might go better for oneself if one has strong desires and dispositions, even strong enough to prevail sometimes over self-interest. In this chapter I shall ask whether benevolent concerns are among these.

First, what exactly is benevolence? According to Brandt's careful analysis of the concept, benevolence involves both affective and motivational components.¹ I have a benevolent concern for you if (a) I find it intrinsically pleasing to learn that you are prospering, and intrinsically distressing to learn that you are suffering, and (b) I am intrinsically motivated (to some effective extent) to promote your good.

The prudential costs of being benevolent are tolerably obvious. First, in so far as one is concerned about the welfare of people whose suffering one cannot relieve, this concern will cause one distress. Indeed, there do seem to be people who are so oppressed by their knowledge of the amount of suffering in the world that they themselves cannot be happy.² Second, strong concern for the welfare of others can of course often lead one to do what is worse for oneself.

This last claim seems as obvious a truth as any. But, since the Desire-fulfilment Theory of self-interest holds that, to put it very roughly, I do what is best for myself by

fulfilling my strongest desires, at least some versions of the Desire-fulfilment Theory of self-interest seem to be saddled with the claim that, if my strongest persisting desire is an altruistic one that your pain be relieved, I can't really make myself worse off by relieving your pain. That is, some forms of Desire-fulfilment Theory seem to make intentional self-sacrifice conceptually impossible.

But consider again the case where my strongest informed desire is to relieve your pain. Now suppose that the fulfilment of this desire has the further consequence that I suffer intense pain everyday for the rest of my life. It is just not plausible to claim that in these circumstances I have done what is best for myself in bringing about the fulfilment of my strongest desire. Indeed, any version of the Desire-fulfilment Theory which does not allow conceptual room for acts which fulfil the agent's strongest persisting informed desire and yet are acts of self-sacrifice needs to be revised so that it does allow conceptual room for this phenomenon.³ So I shall take it that acting on a benevolent desire, even if it is one's strongest persisting desire, can be against one's own good. (Note that I write 'can be', not 'is': I of course recognize, as does every theory of self-interest I mention in this thesis, that doing what will fulfil one's benevolent desires can be, and often is, in the agent's own best interest. I shall come back to this point later.)

There is another prudential cost associated, not actually with having benevolent concerns, but with developing them. There may be transition costs attendant on developing benevolence,

e.g., effort. If there are such costs, they might tip the prudential scales against embarking on a process of becoming benevolent. More dramatically, some people may be completely unable to develop benevolence. I don't know whether there are any such people. But if there are, their trying to develop benevolence is certainly imprudent.

SECTION 2

But let us focus on people who have already or could now develop benevolent desires. Are these people better off with, rather than without, some benevolent desires? It is a commonplace that virtually everyone desires concern and affection from some others.⁴ (Hermits and psychopaths are possible exceptions. If they are exceptions, the following argument would have to be qualified to allow for them. But in order to facilitate discussion I shall for the most part ignore these complications.⁵) We might even say that the desire for concern and affection is a psychological need, meaning that it is a particularly strong and persistent desire that is innate or develops naturally, given human nature and more or less normal circumstances, and that once this desire is established, expunging it would be extremely difficult. Let us now add that it is much more likely that others will have affectionate feelings for a person if they think he or she has a sincere concern for them. Therefore, the psychological need for others' affection and concern would seem to constitute a self-interested consideration in favour of having benevolent concern for them.

But does this consideration favour merely appearing

concerned about others, or is it genuine concern that is prudent? The common-sense reply is that our need is precisely for interpersonal relationships of mutual concern.⁶ And such relationships are logically impossible unless both really have concern for the other.

But even if all we needed was the concern of others, not necessarily the concern of others we cared about, there would still be the point that others are much more likely to care about us if they think we care about them. Now maintaining indefinitely the false appearance of benevolence towards others is exceedingly difficult.⁷ And when A finds out that B's apparent benevolence was and is merely an insidious pretence, A's concern for B will probably vanish, or at least fade.

We must be careful not to overstate what follows from these claims. What follows is not that it is prudent to be strongly and equally concerned about the interests of every other person. Nor does it follow that prudence requires one to be concerned to some degree about every other person. What does follow is much more modest: one has prudential reason to have concern for whomever one needs to have concern for oneself.

An upshot of this is that prudence recommends to different people different degrees of benevolence. I may need concern and affection from many others, in which case it may be prudent for me to have benevolent concern for many people. You, on the other hand, may not need concern and affection from many people, in which case it may be prudent for you to have benevolent concern for fewer people.

SECTION 3

I have now discussed what seems to be the best prudential argument in favour of developing or maintaining benevolent desires. But there may be other prudential arguments to consider. For example, there is the idea that having these benevolent feelings or dispositions and acting on them yields internal rewards.

But what are these internal rewards? Suppose they are supposed to be pleasures. So the idea would be that the agent will get more pleasure out of life if he or she has (or comes to have) benevolent concern for others. In support of that idea, someone might claim that the gratification of benevolent desires is particularly pleasurable.

In order to assess this idea I need to make a digression into the concept of pleasure. According to one account, pleasure is a distinctive kind of feeling or sensation. If we accept this account, the claim that the satisfaction of my benevolent desires is particularly pleasurable means that I get more of a particular kind of sensation or quality of sensation when my benevolent desires are gratified than when other of my desires are gratified. But this first account of pleasure seems to require that there is one sensation common to all pleasures, that there is the same sensation in refined emotional and intellectual satisfactions as in sensual ones. There is, however, no one sensation that is common to acting morally and all the other ways of acting generally taken to yield pleasure.⁸ So this first account of pleasure seems to be on the wrong track.

A second account holds that there is no one special

feeling common to all the sensations or experiences we call pleasant but rather a family of different sensations each member of which we call pleasant.⁹ This account is dogged by *two* objections. The first is that the account does not say why the term 'is pleasant' is used of just this set of experiences. But this objection may not be compelling, because someone might reply that the different experiences bear a 'family-resemblance' to one another, or perhaps that the term 'is pleasant' has more than one meaning, so that it applies to different things in different contexts. The second objection to this account of pleasure is, however, much more powerful. While this account holds that the pleasantness and unpleasantness of experiences could not be distinguished from the sensations of those experiences, apparently there are chemical or surgical techniques which can take away the aversiveness of particular sensations without changing the sensations themselves.¹⁰ And we can imagine the same happening for the attractiveness of the sensations normally found pleasant.

Such examples bring out the conflict between the idea that 'pleasure' is a certain sensation or set of sensations and the idea that, necessarily, if an experience is pleasant for someone, he or she prefers that experience. Now, which of the two ideas one gives up will probably depend on one's theoretical motivations or perhaps on one's linguistic intuitions. But since our theoretical motivation here is to give an account of the sense in which getting pleasure (other things being equal) makes one's life go better in self-interested terms, we had better construe 'pleasure' in the way that links pleasure with desired

experience.¹¹

On this analysis of pleasure, the claim that the internal reward of acting benevolently is sweeter than the internal reward of acting on some other desire amounts to nothing more than the claim that the agent prefers the mental state involved in acting benevolently to the mental state involved in acting on the other desire. But, with this account of pleasure in hand, we must notice that if the agent had developed non-benevolent desires of the same strength, the pleasures arising from the fulfilment of those desires might have been just as pleasing. In fact, cruel people often are able to take delight in the fulfilment of their malevolent desires. And, on the face of it, their revelry in the suffering of others could afford them just as much pleasure as benevolent people receive from watching or thinking about the flourishing of others. Thus, if a prudential argument for being benevolent comes from internal rewards, it is not one based on the quality of the mental states one gets from being benevolent.

A Desire-fulfilment Theorist would say that if we are to construct an argument based on internal rewards, these rewards should be construed not as pleasures necessarily, but rather as the fulfilment of desires. This theorist might then claim that we can easily develop benevolent desires, and that this is a prudential reason for trying to develop them rather than trying to develop some other desires. But is it so easy to develop benevolence? Most people would presumably find it fairly easy to develop benevolent concern for the welfare of those who are in some way closely connected to themselves, such as their family, neighbors, colleagues, and the like,¹² but quite difficult to

develop real concern for those with whom they share little beyond a common humanity.

Here we see a measure of convergence between the Desire-fulfilment Theorist and the Hedonist. The Desire-fulfilment Theorist says that it is prudent to try to develop benevolent desires towards those closely connected with us, because these desires are relatively easy to develop. The Hedonist says that benevolent concern directed towards those closely connected with us is more likely to bring us pleasure than such concern directed towards others. As I have already hinted, the fulfilment of our benevolent desires affords us pleasure just when the prosperity of the people we care about impinges (more or less vividly) on our awareness. Now the prosperity of those with whom we come into regular contact is more likely to impinge on our consciousness than the prosperity of others. We are thus more likely to be able to get pleasure from benevolent concern for these people. So the Hedonist would conclude that, in so far as it is prudent to develop benevolent concern for others, the others towards whom it is generally most prudent to develop benevolence are those with whom we are closely connected.

Now let us consider ^{what} an Objective List Theorist might say about the internal reward of having and acting on benevolent desires. First, according to most versions of the Objective List Theory, one of the things on the list may well be the having of some interpersonal relationships of mutual concern and affection. (Plausible versions do not say that one of the things on the list is having relationships of mutual concern and

affection with everyone.) Obviously, having a relationship of this kind with someone necessarily involves having benevolent concern for him or her. So having benevolent concern for some others is a part of one of the things on the Objective List. Second, pleasure is one of the items listed by any plausible Objective List Theory, and the having of benevolent desires towards certain people is, as we saw above, likely to give one pleasure. Thus Objective List Theories might say that there are two internal rewards of having benevolent desires: having these desires is part of having interpersonal relationships of a certain kind, and having these desires can bring one pleasure. Notice again that the internal rewards, as they are construed by the Objective List Theorist, provide a prudential reason for having benevolent concern for some, but not all, others.

SECTION 4

Are there any other self-interested considerations in favour of being benevolent? There are if in one's society social penalties are imposed on people who are thought to be insufficiently benevolent, or social rewards bestowed on those thought to be particularly benevolent. In the following chapter I shall explore the notion that it is prudent to obey certain rules backed up by legal or social sanctions. There, I shall be concerned with rules such as those prohibiting assault, robbery, promise-breaking, etc. The point being made here, by contrast, is that a society might require—over and above the observance of rules against assault, robbery, promise-breaking, etc.—a degree of benevolence. If society requires from its members some

benevolence (in addition to compliance with such rules) and this requirement is backed up with social sanctions, then these sanctions would operate as a prudential consideration in favour of being benevolent to the extent expected. Of course, different societies might expect different levels of benevolence from their members. So, again, the level of benevolence that is prudent for some people might not be prudent for others.

Is there any other prudential consideration in favour of being benevolent? Well, we benefit from various kinds of co-operation from others. Furthermore, it might be said, this co-operation will probably not be forthcoming unless those others have some friendly feelings for us, and they will not have these friendly feelings towards us if they perceive that we do not have the same for them. These claims are supposed to support the conclusion that there is a self-interested consideration in favour of having some benevolent concern for prospective co-operators.

But let us reconsider the premisses of this argument. We can certainly agree, first, that everyone needs at least some forms of co-operation from others, and, second, that mutual concern and affection often breed or sustain co-operation. (We can add that the causal relation probably works in the other direction as well.) And we must grant that, under conditions in which liking others is a necessary condition of securing from them the kinds of co-operation one needs, it is in one's interest to develop this benevolent concern for those people. On the other hand, as many have argued,¹³ since others need our co-operation just as we need theirs, co-operation can get started

without friendship. Usually, we can get others to co-operate with us in mutually beneficial practices by merely making our co-operating with them contingent on their co-operating with us; having concern for them is not a prerequisite. So the argument which says that our need for others' co-operation generates a prudential consideration in favour of having benevolent concern for them will have only a limited application.

In this chapter I have discussed four possible prudential considerations in favour of having benevolent desires towards some others. These were, in decreasing order of importance: (1) because having benevolent concern for some others is needed in order to get their concern and affection; (2) because it produces internal rewards, as these would be variously construed on the different accounts of self-interest; (3) because there might be social sanctions requiring or at least rewarding a certain level of benevolence; and (4) because friendly concern for others may sometimes be necessary in order to secure their co-operation. Of course, there may be the some people who don't need the concern of others, who wouldn't get sufficient internal rewards from having benevolent desires, who live in a place where there are no social sanctions encouraging benevolence or who need not take notice of these sanctions, and who can get others to co-operate with them without having any benevolent concern for these others. But people of whom all these things are true are probably very rare. For the rest of us, the four prudential considerations combine to make it prudent, on the whole, to have benevolent concern for some others. But, I emphasize, this is

not to say that having a generalized benevolence is in one's own interest. My conclusion is instead that a narrowly focused benevolence is all that is likely to be prudent (unless one lives in a society in which social sanctions strongly encourage wider benevolence).

But, according to all plausible moral theories, being moral involves more than having such a narrowly focused benevolence. Some of the worst villains, we are told, have strong feelings of benevolence and affection, at least towards a few others. These people may be benevolent to the degree I have argued is prudent, but they are not moral. So, to conclude that it is in one's self-interest to have benevolent concern for some others is not yet to conclude that it is in one's self-interest to be fully moral.

CHAPTER 5

SELF-INTEREST AND OTHER MORAL DISPOSITIONS

SECTION 1

What does one need in addition to benevolence, especially narrowly focused benevolence, in order to be a moral person? One view is that a moral person has a number of different dispositions—e.g., to avoid injuring others, not to steal or damage other's property, to tell the truth, to keep promises and agreements, to rescue others when the need arises, to be grateful to benefactors. Another view is that a moral person has a single moral disposition—to do what he or she thinks morality requires. In this chapter, I assess arguments which might be thought to show that living the life that is best for oneself coincides with being moral, where being moral is taken to involve having either the plural dispositions or the single disposition to do what one thinks morality requires. In this discussion I want to remain neutral between the plural disposition view and the single disposition view. To cut down on tiresome repetition, however, I shall usually talk in terms of only one or the other view. But on every such occasion, substituting the other view should not change the truth value of my claims.

SECTION 2

I shall first discuss the idea that an Objective List Theory of self-interest would hold that being moral is prudent. For the sake of argument, I shall not worry here about whether the Objective List Theory is the best account of self-interest. The

focus is instead on the question of whether, given an Objective List Theory of self-interest, being moral is best for oneself.

What argument could there be that this theory of self-interest would hold morality to be prudent? Such an argument might start from the premiss that a plausible sort of Objective List Theory will say that achievement (accomplishing things of real worth) is one of the central values that can make one's life go well. Then, to that premiss, someone might add the premiss that living a moral life is an achievement in the relevant sense. Someone might say that from these two premisses it follows that living morally is, in itself, prudent.

But let us look more closely at this argument. Even if achievement is on the list, it is only one of the things on the list. Similarly, even if living morally is an achievement in the relevant sense, it is only one such kind of achievement. Thus, even if living morally does indeed constitute some benefit to the agent, living morally might also bring such large losses in terms of the other values on the list that, on the whole, the agent might be better off not living morally. In other words, what can validly be concluded from the argument in the previous paragraph is merely that living morally constitutes a benefit to the agent.

But is even this weaker conclusion correct? No doubt being moral benefits the agent in cases in which he or she derives pleasure from being moral. (I shall come back later in this chapter to the matter of deriving pleasure from being moral.) And being moral may be a means to other benefits, e.g. the co-operation of other people. (I shall return later to this

also.) But the issue here is whether the most plausible Objective List Theory holds that being moral is, in itself, beneficial to the agent, i.e., beneficial apart from any pleasure or other benefits to which it may be a means. In favour of the view that it is, someone might say: 'given that writing a great novel is an achievement that, in itself, makes one's life go better in self-interested terms, why wouldn't righting a great wrong also be such an achievement?'

I do not believe that righting the great wrong would, apart from whatever pleasure or social rewards one got out of it, make one better off. Compare the lives of two people, A and B. A's life has been wholly lacking in achievement, but it has been full of deep and long-lasting friendships. Consider now B's life. B has led a very moral life—indeed, B's entire adult life has been spent righting some great wrong. But B's life has been devoid of pleasure or comfort, friendship, knowledge, achievement (other than his moral successes) or any of the other things on the Objective List. Now I want to ask, Do A and B deserve equal amounts of sympathy? A deserves some sympathy because his life has contained no achievement, but his claim on our sympathy is mitigated by the fact that his life contained much in the way of friendship. In contrast, the fact that B led a moral life in no way mitigates B's claim to our sympathy. If B had led a less moral life, he would not deserve more sympathy. It thus appears to be false that being moral is, in itself, (i.e., apart from whatever pleasure or social rewards it brings) a self-interested good.

SECTION 3

I turn now to the idea that, since one is better off if everyone has moral dispositions than one is if everyone does not, it is prudent to be moral. This thought might seem to indicate that being moral is in one's own self-interest. Hobbes observed that if everyone lacks certain inhibitions and society lacks any means of enforcing compliance with certain rules, everyone will be worse off. Such a society would be miserable for everyone, even for the strong, because even they are vulnerable when they are sleeping. Thus, as Kurt Baier puts it, 'universal obedience to certain rules overriding self-interest would produce a state of affairs which serves everyone's interest much better than his unaided pursuit of it in a state where everyone does the same.'¹ So, this argument concludes, if others will not be moral unless I am, prudence tells me to be moral.

But does this mean that the Hobbesian argument shows that self-interest would tell me to develop in myself moral dispositions? No, the situation that is best for me in self-interested terms is that in which I can enjoy the benefits of there being practices promoting personal security, co-operation, mutual trust, and the rest, without having to conform with the practices on the occasions when such conformity would not be advantageous to me.² Consider, for example, the advantages of a situation in which I can depend on everyone else to keep their promises but I do not have the inconvenience of having to keep mine.

The Hobbesian reply will be that unless everyone

(including oneself) behaves morally, morality will be undermined. This is usually incorrect: the behaviour of any one person will hardly ever make a critical difference to whether moral practices endure or die out. This is because most people's commitment to morality is not contingent on everyone else's being moral: they know that there are plenty of free-riders and scoundrels around, but they continue to be moral.³ Furthermore, people may be able to gain access to the benefits of many moral practices (e.g. the practice of truth-telling) by merely appearing to be observing the practices. So one person may not lose out on the benefits of these practices if his or her non-compliance is secret. (I shall return to the possibility of secret non-compliance in a moment.) We can see, then, that the Hobbesian premiss 'Unless one is moral others won't be' is ordinarily false. So the Hobbesian argument fails to prove that being moral is in one's interest.⁴

SECTION 4

Someone sympathetic to the Hobbesian line of thought might now move to the claim that it is external sanctions that make being moral prudent. As I said in Chapter 2, by 'external sanctions', I mean legal sanctions, such as capital punishment, imprisonment, or fines, and social sanctions, such as distrust, resentment, the withdrawal of goodwill, or ostracism. People usually distrust, ostracize, or even take legal action against whoever they believe has committed seriously immoral acts. It is supposed to follow that since we desire not to be distrusted or ostracized—not to mention fined, imprisoned, or

executed—being moral serves our interests.

But there are a number of problems with this argument. The first is that external sanctions do not necessarily attach either to the best justified morality or to the moral code the agent accepts. Instead, they attach to whatever moral code the society accepts. In other words, they back up the positive morality. Now the particular positive morality might be grossly unjust and inhumane—i.e., it might be a long way short of anything that could correctly be understood as the best justified morality. Since this is true, it is false that unless one behaves in accordance with the best justified morality (or even with what one believes to be the best justified morality) one will suffer external sanctions. On the contrary, since the positive morality may enjoin actions that would be forbidden by the best justified morality, the external sanctions can provide prudential reasons for doing such actions. This is a point of the first importance. For it leads to the conclusion that all we can accomplish by appealing to external sanctions is to show that conformity with the positive morality is prudent. It is a further (and large) question whether the positive morality happens to coincide with the best justified morality.

Furthermore, even if we accepted the basic principles of the positive morality, we might disagree with the rest of the society about what particular concrete actions follow from the positive morality. Suppose I am the member of a society the positive morality of which permits the publication of any true information about the government as long as the information is not inimical to national security. And suppose I know that the

publication of certain true information about the government would not be inimical to national security. I could then conclude that the principles of the positive morality would not in fact forbid the publication of my information. But suppose most people wrongly believe that the publication of any information about the government would be a risk to national security, and that therefore the publication of my information about the government would be immoral. To stay clear of the society's negative external sanctions I must not publish my information, even if I know that the rest of my society is confused about what the positive morality would forbid. The general point this example illustrates is that the external sanctions provide a prudential consideration in favour of doing what the the society believes follows from its moral code, even if this belief is mistaken.

And there are further problems with the idea that external sanctions make it imprudent to be immoral. Even if it were never prudent to do openly an act one's society thinks its morality forbids, this would not entail that prudence requires actual conformity with the positive moral code. Maybe the mere appearance of conformity is enough.⁵

Moreover, taking into account only external sanctions, we must grant that even a course of action that involves openly infringing the positive moral code can sometimes maximize the agent's own good. Sometimes a known wrongdoer is not punished by the law. This might occur because there is insufficient legally admissible evidence to get a conviction in court. Or it might occur because the person infringed a norm of the positive

morality that was not reflected in the enforced statutes of the legal code. But whatever the reason the wrongdoer is not punished by the law, the negative social sanctions (distrust and ostracism) might not be large enough on their own to offset the prudential gains. And there may be cases in which people do something for which they suffer legal punishment as well as some social censure and yet still come out better off as a result of their crimes. A robber who spends a few years in prison and then goes off to enjoy the stolen millions may be an example.

That last example raises a further problem. What makes that example particularly plausible are reports of criminals who emigrate to a different society in which they receive no social censure for their previous actions. This can happen because what is condemned by people in one part of the world differs from what people in other parts condemn. But it is also true, as I already suggested in Chapter 2, that different groups within the same society have somewhat different moral principles, though it might also be true that too great a divergence would cause the society's breakdown. (This may make it difficult to determine just what the positive morality is. But, as I announced in Chapter 2, I shall use the term to refer to the principles and norms backed up by the social sanctions deployed by the society's majority.) So the attitudes of an individual's own group might favour somewhat different behaviour than that demanded by the majority's moral code, as well as by the law. This fact sometimes figures in the explanation of how someone known to have done something against the positive morality (the morality of the society's majority) can nevertheless come out better off for

having done it.

All I have done so far is to point out (a) that, no matter how iniquitous and evil the positive moral code is, it is this code that external sanctions back up; (b) that no matter how confused the society is about what follows from that code, the prudential considerations yielded by external sanctions favour doing the particular actions the society believes its code demands; (c) that external sanctions directly enforce only the appearance of doing these particular acts; and finally (d) that even when negative external sanctions are deployed against someone known to have violated the positive morality, the costs the perpetrator thus incurs are sometimes outweighed by the prudential benefits of that violation. These points must all find their place as qualifications to the claim that external sanctions make being moral prudent.

On the other hand, in the vast majority of cases, a person's open violation of the more serious principles of the positive morality would have disastrous results for him or her. So I shall henceforth ignore the qualifications necessary because it occasionally is ^{false} ~~is~~. Likewise, I shall abstract from the difficulties raised by the fact that people may disagree about the first principles of morality and about what particular acts are enjoined by the positive morality. But what is far too important to set aside is the point that the external sanctions attach to the positive morality, not necessarily to the best justified morality.

SECTION 5

Because of the likelihood that they will suffer external sanctions as a consequence of overt infringement of the positive morality, most prospective immoralists would attempt only covert infringement. Of course, infringements that are meant to be secret might get exposed. The agent may, however, know the probability of exposure and then be able to choose accordingly. On the other hand, the agent may not know the probability of exposure. It might then be argued that our general lack of knowledge about the probabilities of being caught furnishes a good prudential argument for generally being moral. In this section I shall explore this suggestion.

As I mentioned earlier, the net expected prudential value of an act is calculated by multiplying the prudential gains which may result from that act by the probability that those gains will occur, and multiplying the prudential losses that may result from the act by the probability that they will occur, and then subtracting the expected loss from the expected gain. It may, of course, be objected that this picture is ludicrously unrealistic in that it supposes that the variables can be quantified, which is rarely the case in practice. Often we have only rough estimates, and to ask that such estimates be precisely quantified would seem absurd. I accept this point. But in this section, because I want to concentrate on problems raised by probability, I shall set aside the difficulty of quantifying possible gains and losses. I shall, however, consider unquantified probability assignments later in this section.

As I said, the agent may or may not know the

probabilities of the various possible outcomes. Let us call choices made when the agent does not know the probabilities, 'choices under uncertainty'; and choices made when the agent does know the probabilities, 'choices under risk'.⁶

Obviously, in choices under risk the agent can know which act has the greatest net expected prudential utility. Furthermore, the agent can know that this act is an immoral one. A bank clerk, for example, might have reliable statistics about how often bank clerks get away with embezzling money from their employers.⁷ The probability of being caught and punished might be tiny and the personal gain from the embezzled money might be enormous. Then, taking into account only external sanctions, we cannot deny that the choice with the greatest expected benefit to this bank clerk could be to embezzle the money.

In choices under uncertainty, one does not know the probabilities and so cannot calculate the expected values of alternative acts. It might seem that, under these conditions, some sort of 'maximin' strategy, i.e. opting for the best worst possible outcome, is rational.⁸ According to this line of thought, since being discovered to have committed some serious violation of the positive morality usually has calamitous consequences, when choosing under uncertainty the agent would be prudent not to violate the code.

But there are objections to the use of maximin as a prudential decision strategy for use when choosing between outcomes ranked in terms of prudential value (or as we might call it, personal utility). Maximin does have considerable appeal

when the alternative possible outcomes are judged in terms of resources (e.g., money). Usually it is the case with resources that the more one has of the resource, the less equally sized increments matter. In other words, resources normally have declining marginal utility. Utility itself, however, never does. So maximin seems less appropriate when outcomes are ranked in terms of personal utility than when they are ranked in terms of resources.

Other objections to maximin take the form of elaborations of alternatives to maximin. For example, it has been argued that, when one is in a state of complete ignorance about the probabilities of the possible outcomes (but does know what outcomes are possible) then the rational strategy is to take account of only the best and the worst possible outcomes.⁹ Another approach is to treat choices under uncertainty on the model of choices under risk, thereby doing away with the rationale for maximin as an alternative to maximizing expected prudential value. In order to treat choices under uncertainty as if they were choices under risk, the principle of insufficient reason is brought in to underwrite the assignment of an equal probability to every possible outcome. The assignment of equal probabilities is held to be justified on the grounds that there is no reason for thinking any possible outcome less likely than any other. But this approach seems suspect. In choices under uncertainty, the agent has no information about the probabilities, and so the agent has no more reason for supposing that they are equal than for supposing that they are unequal.

What, then, is the best method for making, under

conditions of uncertainty, the choice that will turn out to maximize one's own good? One might try to postpone the decision until one acquires some information about the probabilities. Of course, this may not be possible. In cases when it isn't possible, maybe the best method of maximizing one's own good would be to choose arbitrarily and without delay, since time taken in worrying is a prudential cost.

But much more important than the question of how we should decide when we have no probability information is the point that only rarely do we have to make decisions under conditions in which we have absolutely no clues about the probabilities of the possible outcomes. Complete ignorance about probabilities and complete certainty about probabilities are opposite endpoints of a spectrum, and usually choice conditions lie somewhere between these endpoints. That is, most choices are made under conditions of partial uncertainty. Often we have some information about probabilities and yet doubt the reliability of this information. Or we have some evidence for particular probability assignments and yet also believe that this evidence is not conclusive. Obviously, the strength of the evidence or the reliability of the information is sometimes more and sometimes less.¹⁰

How should the agent deal with this complication? In order to answer this question, let us first consider a case where the agent is choosing between two acts each of which have two alternative possible outcomes. And let us suppose the prudential values of the two possible outcomes of the first act are 100 and 0 respectively, and the prudential values of the two possible

outcomes of the other act are also 100 and 0 respectively. Of course, it is proper to let the probabilities decide in such a case. If the probability of the good outcome of the first act is high, and the probability of the good outcome of the second act is low, then prudence tells the agent to choose the first act.

But now consider a second case. This second case is like the first case in that the prudential values of the two possible outcomes of each alternative choice are the same (i.e., 100 and 0). But in this second case, the information the agent has indicates that the probabilities of the outcomes of the alternative acts are parallel. For example, if I bet on horse A, I get 100 if A wins but nothing if any other horse wins, and the same, mutatis mutandis, with horse B. And I have been given a tip from Alfred that the likelihood of A winning is 60% and a tip from Brian that the likelihood of B winning is, again, 60%. Now, if Brian is more reliable than Alfred, I should of course follow Brian's advice and bet on B. To put the point in a more general form: the agent should let the reliability of the information about the probabilities decide the issue when all other things are equal.

And why shouldn't the reliability play a part even when all other things are not equal? Consider a case in which the possible payoffs are the same as above, and again Brian tells me the probability of B winning is 60%. But this time Alfred tells me that the probability of A winning is 80%. Of course if Alfred and Brian are equally reliable, the prudent choice is to bet on A; but suppose Alfred is only 30% reliable, while Brian is 90% reliable. Multiplying the variables by one another, we get (for

the bet on A): $100 \times 80\% \times 30\%$; and (for the bet on B): $100 \times 60\% \times 90\%$. That is, 24 -vs- 54. But is this the right method? My own guess is that it is—I know of no good reason for giving the reliability of the information about the probabilities of the possible outcomes less (or more) weight than the other variables.

As I have already acknowledged, we are rarely able to quantify all the relevant factors. But that does not undermine my point. Even if we can't say anything more precise than that the information supporting one probability figure is somewhat more reliable than the information supporting another, this unquantified difference can still rationally influence the prudent agent to choose one act rather than another. So the existence of uncertainty does not necessarily make being moral prudent.

SECTION 6

At this stage in the discussion, three points might be made by those claiming that external sanctions make it prudent to have an overriding disposition to conform with the positive morality (or at least its more serious principles and norms). The first starts from the observation that the prudential reasoning of many of us is routinely distorted by certain biases, such as the bias towards the nearer future.¹¹ This is a bias towards outcomes in which the possible benefits come sooner and the possible costs come later (often the case with violations of positive morality). Of course, the bias towards the nearer future is not a mistaken belief but a matter of differential concern; nevertheless, like other kinds of differential concern, it can induce one to

make miscalculations. It can lead one to overestimate (1) the prudential benefit of possible outcomes of a prospective immoral act, or (2) the probability that these outcomes will in fact be realized, or—what is most relevant here—(3) the reliability of information that the probability is high that these outcomes will be realized. Likewise, it can lead one to underestimate (1) the losses that would result from being caught, (2) the probability of being caught, and (3) the reliability of information that the probability is high that one will be caught. In Hume's words, '[knaves] purpose to cheat with moderation and secrecy, a tempting incident occurs, nature is frail, and they give into the snare'.¹² Therefore, if our infringing the positive morality has a possible outcome that would benefit us in the near future, and our prudential calculations come out favouring this act, then we should be suspicious of these calculations.

Of course, trying to protect ourselves against our weaknesses is not our only option—we might try to get rid of them. Thus we might take steps to eliminate the bias towards the nearer future and its distorting effects from our prudential reasoning. Perhaps there are some people who could do this. For such people, the prudent course of action might be to make more accurate their prudential calculations, and then to trust such calculations even when they indicate that the act with the greatest expected prudential value is an act condemned by the positive morality. So the first point I conceded—that our prudential reasoning may be distorted—does not entail that external sanctions make it prudent to have an overriding disposition to be moral.

Turn now to the second point likely to be made by those

who argue that external sanctions make it prudent to have this overriding disposition. This is the point that one is less likely to suffer external sanctions at some point in one's life if one has a overriding disposition to conform always with the positive morality than if one has a disposition to conform with it except when one thinks one could get away with not doing so. Presumably, there is a rough correlation between the number of times one infringes the positive morality and the likelihood that one will at one time or another feel the bite of external sanctions. Indeed, it is probably very rare that someone gets through life infringing the established norms whenever this seems to be in his or her best interests and yet never suffers any external penalty.¹³

But from the fact that one is more likely to suffer external sanctions if one leads such a life, it does not necessarily follow that one will have a less good life in self-interested terms. Let me describe an imaginary example to illustrate this point. Ted led a life of always doing what he believed would be best for him, on the whole, in self-interested terms. During the course of his life he often violated the rules of the moral code current in his society. He could not keep all of his violations secret. But he did get away with many of them, and these had large payoffs, large enough even to compensate him for the losses he suffered on the occasions when he was caught. So Ted's life went better in self-interested terms as a result of his having the overriding disposition to do what he believed was best for him than it would have gone if he had instead had an overriding disposition to comply with the positive morality.

Therefore, the second point I conceded does not prove that external sanctions make it prudent to have an overriding disposition to comply with the positive morality.

Of course, many people do have a lot to lose from being caught committing even a single violation of an important principle of the positive morality.¹⁴ This brings me to the third point to be conceded to those arguing that external sanctions make it prudent to have the disposition to conform always with the positive morality: attempting to get away with some infringement of the positive morality involves often the cost of having to be extremely careful never to give oneself away.¹⁵ This cost must of course be included in one's prudential reasoning. But this cost will not always be large enough to make such acts imprudent. Occasions do arise when one knows that not giving oneself away would not be very taxing. So, apparently, one would be better off from the standpoint of self-interest if one were disposed in such a way that one would take advantage of such opportunities. So, just as with the first two points I conceded, the third point I conceded does not establish that external sanctions make it prudent to have an overriding disposition to conform always with the positive morality.

SECTION 7

Having argued that external sanctions are not on their own up to the task of reconciling morality and self-interest, I want now to ask whether self-interest and morality are reconciled if we broaden our focus to take in internal sanctions. First, what

are internal sanctions? People with moral dispositions typically feel guilty when they fail to do what they are morally disposed to do. Furthermore, there are, it is often thought, internal rewards in the form of pleasant feelings experienced by moral people when they act, or remember acting, morally. For instance, Kavka writes, 'there are special significant pleasures or satisfactions that accompany regular moral action and the practice of a moral way of life which are not available to (unreformed) immoralists and others of their ilk'.¹⁶

It may be worth emphasizing at the beginning of this discussion that, while it is to the rules of positive morality that external (i.e., social and legal) sanctions are attached, internal sanctions operate as prudential considerations in favour of acting in accordance with the agent's own moral dispositions and beliefs, not necessarily with the principles of the positive morality. Since the agent's own moral dispositions might well diverge from the moral principles and norms current in his or her society, the prudential considerations furnished by internal sanctions and those furnished by external sanctions might conflict. (I return to points this raises later.)

I want now to look more closely at the nature of the positive internal sanctions. It follows from the picture I gave of pleasure in Chapter 4 that one experience is more pleasant than another just if the person having the experience prefers the first to the second as an experience. We might now ask whether, on this account of pleasure, people get pleasure from the fulfilment of their moral desires. The answer is, of course, sometimes Yes and sometimes No.

First, I may fail to get pleasure from the fulfilment of one of my moral desires when this desire extinguishes before its fulfilment. Second, even when the desire does not extinguish before its fulfilment, I might never find out about the fulfilment of my moral desire and therefore fail to get any pleasure from it. Third, even in cases where my desire has not extinguished and I know about its fulfilment, I may (more or less consciously) shut out of my mind the fact that my desire is fulfilled. The fulfilment of one's desire, if one has blocked the very desire from one's own awareness, does not provide one with pleasure. Why might I block this desire from out of my mind after acting on it? I might believe that conduct is morally virtuous only if its motives are completely unselfish, and that in order to preclude selfish motives, one should not let oneself benefit in any way from doing one's duty. Thus after I have acted morally, I may concentrate my thoughts on those of my desires that will be frustrated as a result of my moral act and simultaneously put out of mind the moral desire I acted on. Fourth, even if I have no inhibition about paying attention to the fact that my moral desire is fulfilled, the circumstances resulting from my moral act may be such as to keep me from devoting attention to the fact that my moral desire is fulfilled, and this may stand in the way of my getting pleasure from its fulfilment. To take an extreme example, suppose that as a consequence of fulfilling one of my moral desires, I am immediately placed in constant pain or danger which completely usurps my attention. Here I have not really lost my moral desire; it has just that I can't think about its fulfilment,

because other facts are demanding all my attention. In less extreme and more common cases, I would be able to devote a limited attention to the fulfilment of my moral desire and therefore get some pleasure from it, but not as much as would be possible if the circumstances didn't force me to concentrate on something else.

To be sure, people often do get pleasure from the fulfilment of their moral desires.¹⁷ This occurs when the moral desires have not extinguished, the people know about the fulfilment of their desires, and they attend to the fact that their desires are fulfilled.

I shall now explain what I think is the most plausible prudential argument that involves internal sanctions. As I acknowledged in previous sections, the operations of external sanctions normally make complying^{with} the positive morality (or at least^{with} the more serious principles of the positive morality) prudent. On these occasions, one would be even better off if complying were itself a pleasure rather than a chore. And having moral desires to do the acts in question would help make compliance a pleasure. In other words, a possible prudential benefit of having moral desires is the pleasure that having such desires enables one to get from complying with the positive morality, which one normally has to do whether or not one can take pleasure in it. (It may be worth emphasizing that one's moral desires can enable one to get internal rewards from doing what the external sanctions make prudent only if there is convergence between these desires and the positive morality.)

Now it is time to give proper consideration to the

prudential disadvantages of having moral desires and dispositions. There are three. A person with moral desires and dispositions will, at least sometimes, act morally even when he or she knows that another course of action would be better in self-interested terms.¹⁸ That having moral dispositions can lead one to do such acts is a very considerable price to pay for the benefits of having moral dispositions.

This is perhaps the best place to mention a common mistake about the self-interest of someone who has moral desires. It is true that having moral desires, and so being subject to internal sanctions, greatly alters the sums in one's prudential calculations. There is the internal reward that someone with moral desires could get from acting on them; and there is of course the internal punishment, guilt, he or she could suffer as a result of acting on contravening desires. Because of these benefits and costs, the course of action with the greatest expected benefit for the person with moral desires will sometimes differ from the course of action with the greatest expected benefit for the amoral person. Ms. Jones has moral desires and dispositions; Mr. Smith doesn't. Jones would suffer such intense feelings of guilt were she to cheat on her tax return that she would be better off not cheating. Furthermore, she gets some satisfaction out of being honest. Smith, on the other hand, not being susceptible to the internal sanctions, would be better off cheating. Yet there will be cases in which, even for a *person* like Jones who has moral desires, the guilt the person would suffer as a consequence of doing some immoral act would be fairly small in comparison with the advantages the act

would secure. Likewise, the moral satisfaction that would come from not choosing this act would be small in comparison with the act's advantages. That is, even for people prone to internal moral sanctions, the prudent act might not always be the moral one. So, internal sanctions do not make every act required by morality prudent.

Now for the second prudential disadvantage of having moral sentiments: having them may lead one to miss out on certain pleasures foreclosed to the moral person.¹⁹ It may be that a certain evil pleasure can be taken in treachery and ruthlessness. Even savagery may have its satisfactions. But all such pleasures are lost to the agent in so far as he or she is moral.

I have already mentioned the third prudential disadvantage of having moral dispositions. If one has these dispositions, one will feel guilty when one does what one thinks one should not do. This might actually be the most important disadvantage, at least in the light of the following empirical claims. One claim is that not even the most virtuous people manage always to do what their moral dispositions urge them to do, and hence everyone who has moral dispositions experiences guilt sometimes. Furthermore, even for the most virtuous people, perhaps the pains of guilt are greater than the pleasures of virtue—that is, if we add up all the pleasure people get from acting on their moral desires, and then add up all the guilt they suffer as a result of not doing what they believe they ought to do, the guilt outweighs the pleasure. If both of these claims are true, then being susceptible to the internal sanctions is not a

prudential advantage, but rather a disadvantage. But I am somewhat sceptical about the truth of the second claim, at least with regard to most people. Most of us are quite proficient at explaining away and excusing our own transgressions, thereby avoiding or at least minimizing unpleasant feelings of guilt.

Moreover, no matter what the advantages of having moral desires and dispositions, there might be some people with respect to whom no argument appealing to the advantages of internal sanctions is appropriate. In Chapter 4, I mentioned the possibility that there are people who are just unable to develop benevolent desires. Parallel remarks need to be made here about people who are unable to develop in themselves other moral desires. For these people internal sanctions in no way provide a self-interested consideration in favour of developing these desires. I also mentioned earlier the possibility that there are selfish people who could develop benevolence but who are better off not doing so because of the transition costs involved. Again, parallel remarks are in order about the transition from the state of having none of the moral desires and dispositions I have been discussing in this chapter to the state of having them. That is, there might be some people for whom the costs of developing moral inclinations are larger than the likely benefits of having them.

But for the sake of argument, let us confine our attention to people who could develop these desires and who would not have to undergo prohibitive transition costs. And let us assume that the feelings of guilt most people with moral motivations would experience would not be as great as the

feelings of satisfaction. That is, assume the only prudential disadvantages of having moral motivations are (i) the disadvantage of being impelled under certain circumstances to do what is worse for oneself and (ii) the disadvantage of missing out on some wicked pleasures. Now let us ask: Do the advantages of being susceptible to the internal sanctions outweigh these two disadvantages? Here I can only make a speculative generalization. With respect to most people, the advantage of being able to get pleasure from doing acts that are anyway prudent is likely to be greater than these two disadvantages. If that is right, most people would be prudent to have (or develop) desires to do the kinds of act required by the positive morality.

It is a further question, however, whether prudence recommends that these desires be so strong that they always get their way against self-love. It is one thing to claim that my life is likely to go better in self-interested terms if the relative strengths of, on the one hand, (e.g.) my desire to keep my agreements and, on the other, my desire that my life go as well as possible, in self-interested terms, are such that I sometimes am led to keep agreements rather than do what I think would be best for myself. It is another thing to claim that my life is likely to go better if their relative strengths are such that I would never break an agreement in order to do what is best for myself. Is the second, stronger claim true? On the one side is the point that the pleasures attendant on having overwhelmingly strong moral desires may be more intense than the pleasures attendant on having less strong moral desires. On the other side is the point that overwhelmingly strong moral desires

will lead one to act contrary to one's interest more often than less strong moral desires would. So the question is: Would the additional pleasure that comes with having overwhelmingly strong moral desires be great enough to offset the losses incurred through the additional number of imprudent acts? The answer (as Sidgwick notes) seems to be that most people are so constituted that the additional pleasure they would get would not be great enough.²⁰ Thus it is not in one's own interest for the moral desires and dispositions it is prudent to have to be so strong that they always override self-love.

By way of summing up this chapter, let me say why that last point matters. Earlier in this chapter I argued that prudence would tell most people to have intrinsic desires to do the kinds of act required by the positive morality. This is not to say that on every occasion it is in the agent's own interest to act on these desires. Furthermore, these desires are merely to do the kinds of act required by the positive morality, and the positive morality may be quite different from the best justified morality. These two points are enough to create the possibility of a considerable gap between being prudent and being moral (in the sense of strict obedience to the best justified morality). But now what if the principles constituting a particular positive morality could not be improved upon—in other words, what if the positive morality does happen to be the best justified morality? Even under this condition it is not true that prudence tells one to be moral. First of all, there will still be individual occasions on which the morally required act is not as good in self-interested terms as some other act the agent might choose.

But even if we move from the level of individual occasions to the level of general desires, we find conflict between what would best serve self-interest and what would satisfy morality. For being fully moral presumably involves having an overriding commitment to morality, but, if the argument of the preceding paragraph is correct, this is not prudent.

CHAPTER 6

SELF-INTEREST AND MORAL THEORIES

SECTION 1

This chapter addresses the question 'What moral beliefs is it in one's self-interest to accept?' This question is of considerable interest in its own right. But it is also of interest because of its relevance to the question 'Why should I be moral?' In this chapter I shall argue that there are some moral beliefs or, more accurately, theories it is more in one's interest to accept than others. But suppose one of the moral theories it is not in one's interest to accept is nevertheless the correct or best justified theory. Then not only is it (as I concluded in the previous chapter) not in one's interest to have overriding desires to do the kinds of act required by morality, but also it is not even in one's interest to have the correct beliefs about what morality requires.

Though I shall try to determine which moral beliefs it is one's interest to have, I shall not tackle the deep and vexed question of which moral beliefs are correct or best justified. So the conclusion of this chapter will take a conditional form—if the moral beliefs it is in one's interest to have are correct, then things stand as they did at the end of the previous chapter; but if the moral beliefs it is in one's interest to have are not correct, then this chapter serves to widen the gap between self-interest and morality.

SECTION 2

Offhand, it might appear that Egoism is the most prudent moral theory to accept. For Egoism would never forbid one to do what is best on the whole for oneself. But is Egoism a moral theory? Egoism is a one principle theory—its single principle is that one is required to do what is best for oneself. So we could reformulate our question: could this principle constitute (narrow) morality? This question forces us to ask what the criteria are a principle must fulfil in order to be eligible as a principle of (narrow) morality.

In Chapter 2, I outlined the familiar distinction between formal and material conditions on moral principles. To start with the formal conditions, Egoism is ruled out by the most plausible versions of the view that a principle cannot be a moral principle unless it has certain formal features. Take the idea that moral principles must be universalizable in some weak sense.¹ If it is wrong for one person to do a kind of act, then it must be wrong for any other person similarly situated to do that kind of act. Consider two forms of Egoism: (1) First-person Dictatorship Egoism—everyone is required to serve my interests, and (2) Free-rider Egoism—everyone is required to act justly except for myself, if I choose not to. The indexicals in these principles are of course to be read as individual constants rather than as universally quantified individual variables. These principles cannot be stated without using these individual constants. Therefore, they are not universalizable. And, on the view that moral principles must be universalizable, this disqualifies them from being moral

principles.²

A third form of Egoism claims that one is morally permitted to do what is best for oneself no matter what the consequences are for others. This form of Egoism does not fail the formal requirement of universalizability. It can be stated as a universal proposition: 'For all agents A and events E, if E will maximize A's own good, A is morally required to promote E.' Since this form of Egoism is universalizable, let us call it 'Universal Egoism'. But, although it is universalized, it seems not to be the prudent 'morality' to accept. For the reasons discussed in the previous chapter, one is likely to be better off in self-interested terms if one has desires to do the kinds of act required by the positive morality. Now if a society's morality didn't help resolve problems of partial conflict—which Universal Egoism certainly would not—then the society would probably self-destruct, or at least lose out in competition with some rival society whose positive morality did help resolve these problems. So it is extremely likely that the positive morality of one's society is something other than Universal Egoism (unless perhaps divine sanctions play a role in people's beliefs). Therefore, having desires to do what the positive morality requires will not be the same as being disposed to act in accordance with Universal Egoism. There is thus likely to be some psychological tension between, on the one hand, those desires and, on the other, the belief in Universal Egoism. Consequently, if there is some other normative theory that would sit more easily with those desires, it would be more prudent to accept that other theory than to accept Universal Egoism.

SECTION 3

Having dealt with Egoism, I now turn to the relative prudential advantages of accepting a deontological as opposed to a consequentialist morality. As I explained in Chapter 2, deontological moralities contain agent-relative principles forbidding certain kinds of act and enjoining other kinds even when such acts would produce less good consequences, impartially considered. Consequentialism, in contrast, insists that it is wrong not to do what one knows will produce the best possible consequences, impartially considered.

Now it is usually assumed that a deontological morality is less demanding than a consequentialist one. But, if these two kinds of theory are distinguished as I have distinguished them, there is hardly any limit to the number of each kind of theory. The distinction between theories that do contain agent-relative first principles (deontological theories) and theories that contain only agent-neutral first principles (consequentialist theories) is so abstract that the most we can say about the relative demandingness of the two kinds of theory is that, some consequentialist theories will be more demanding than some deontological theories, and perhaps vice versa.

To some, this claim will seem counter-intuitive. Anyone who finds it counter-intuitive probably assumes that no deontological theory is more demanding than any consequentialist theory. But, as I have defined deontological principles, the principle 'Never do what benefits yourself' is a possible (albeit crazy) deontological principle, because it is agent-relative in

form. Now a deontological morality which consisted of this one principle would be more demanding than virtually all consequentialist moralities.

On the other hand, we can ask whether the most common consequentialist theories or the most common deontological theory is favoured by prudence. I take it that Common-sense Morality is the most common deontological theory. As I said in Chapter 2, this theory tells one not to assault, steal, coerce, lie, break one's promise, etc. It also requires one to give special weight to the interests of those connected in certain ways with oneself, and to come to the aid of others if their gain is much larger than the cost to oneself.

Common-sense Morality seems plenty demanding: it requires us (e.g.) to keep a promise, even when this would be worse for us; or to give a benefit to others under certain conditions even when this would involve some cost to us. The most common consequentialist theories, however, are even more demanding. Consider Act-utilitarianism. This theory is usually understood to have an extremely strong principle of aid: we are required to do what will benefit others as long as the benefit given to them is at all larger than the cost to us. Common-sense Morality, as I have noted, also contains a principle of aid: to make sacrifices for the good of others when the loss to us is small in comparison with the benefit to them. But this principle is weaker than the Act-utilitarian principle of aid. For, in contrast to the Act-utilitarian principle, this principle does not require us to make sacrifices for the benefit of others when the benefit to them would not be much larger than the cost to us.

Thus, as Sidgwick observes, '. . . Utilitarianism seems to require a more comprehensive and unceasing subordination of self-interest to the common good [than Common-sense Morality does].'³

On the other hand, it is not true that on every occasion Act-utilitarianism is more demanding than Common-sense Morality. There can be circumstances under which the act favoured by Act-utilitarianism would, on the one hand, be best for the agent and, on the other, transgress against Common-sense Morality. An example might be the robbery of a well-off person by a badly-off person. For another example, suppose there is some lie Steve could tell Sue which would benefit him more than it would harm her. Such an act would (it seems) not be wrong according to standard Act-utilitarianism, but it would be wrong according to Common-sense Morality. However, though such examples do exist, Common-sense Morality is less demanding in most circumstances (particularly on the well-off, as I shall discuss below).

As I mentioned in Chapter 2, there is a second traditional act-consequentialist theory. This theory holds that the morally right act is that which produces the best consequences judged in terms of welfare and equality. I shall call this theory 'Distribution-sensitive Consequentialism'. Now would Distribution-sensitive Consequentialism be more demanding than Common-sense Morality?

Distribution-sensitive Consequentialism's principle of aid would presumably be: make sacrifices for the benefit of others whenever this will produce the best possible consequences,

judged in terms of welfare and equality. There will be very many cases when, on the one hand, Distribution-sensitive Consequentialism requires one to make a sacrifice for the benefit of someone who is worse off, because this would increase net welfare and equality, and yet, on the other hand, Common-sense Morality would not require this act, because the loss to oneself involved in making the sacrifice would not be very much smaller than the benefit to the other person. To be sure, there are circumstances in which Distribution-sensitive Consequentialism is less demanding than Common-sense Morality. As an example, take a case where I am badly off and you are well off, and yet I could make some sacrifice that would benefit you much more than it would cost me. Common-sense Morality would require me to make this sacrifice, but Distribution-sensitive Consequentialism would not, because my making the sacrifice for your benefit would increase inequality. Usually, however, Distribution-sensitive Consequentialism, like Act-utilitarianism, will be more rather than less demanding than Common-sense Morality.

It is particularly with respect to the well-off that our two kinds of consequentialism are more demanding than Common-sense Morality. Obviously, the more fortunate someone is, the more it is in his or her interests to have a morality in which the duty to give aid to the less fortunate is relatively weak. As Harman points out, the rich and powerful have little to gain by the adoption of a strong principle of aid, because 'given their resources, they would be called upon to do most of the helping and would have little need for such help for

themselves.⁴ Since it is to the advantage of the well-off to have weaker principles about aid to the less fortunate, they would generally be better off with Common-sense Morality than with either of the two main act-consequentialist theories.

In the previous chapter I argued that the closest we can plausibly get to the conclusion that being moral is prudent is the much weaker conclusion that, for most people, it is prudent to have desires and dispositions to do the kinds of act required by the positive morality. Thus the question of which moral theory it is prudent to accept seems to turn mostly on the question of which moral theory is embodied in the positive morality of one's society. If my moral beliefs do not coincide with the principles of the positive morality, then my moral dispositions will probably also fail to be in line with the positive morality. If this is the case, then my moral dispositions and the society's principles will sometimes pull in different directions. And, when they do pull in different directions, I am faced with the choice between two unappealing alternatives. If I follow the social morality, I shall experience guilt as a result of having gone against my own moral dispositions.⁵ Alternatively, if I act on my own moral beliefs and dispositions, I risk incurring external sanctions. I am better off if my dispositions and beliefs are such that these dilemmas never arise. However, as has been said, the moral code entrenched in a society may be repugnantly unfair and inhumane. But individuals in that society nevertheless have a strong self-interested reason for accepting whatever moral theory is embodied in that positive moral code.

Now what moral theory is embodied in the positive morality of our society? Common-sense Morality, as its name suggests, seems to be the answer.⁶ So, each of us would be better off accepting Common-sense Morality than consequentialism. When this point is added to the earlier point that Common-sense Morality is generally less demanding than consequentialism, it seems overwhelmingly clear that, for a member of our society, accepting Common-sense Morality is better, prudentially, than accepting consequentialism.

At this point some consequentialists might object that the preceding discussion has traded on a mere caricature of consequentialism. They might say that there is much less divergence between their theory and Common-sense Morality than I have suggested, and that a rough approximation of their consequentialist theory might indeed be embodied in our positive morality. They might appeal to a form of consequentialism that is Collective and has a multi-level structure, as I explained these features in Chapter 2. Bringing in the multi-level structure shifts us from the question 'Which act should I do in order to produce the best consequences, impartially considered?' to the question 'Which dispositions should I have in order to produce the best consequences, impartially considered?' As I explained in Chapter 2, the consequentialist can say that I should have dispositions against assaulting, stealing, blackmailing, breaking promises, lying, and so on. In saying this, consequentialism does seem to converge with Common-sense Morality (and with positive morality).

But it might be objected that consequentialism, even if

it does have a multi-level structure, doesn't coincide with Common-sense Morality (or underlie our positive morality) because consequentialism demands such extreme impartiality. The consequentialist seems to tell me to have that pattern of concern my having of which will bring about the best consequences, given the small amount of concern most people in fact have for all others except for those connected with them in special ways. What would bring about the best consequences, given the selfishness of others, is for me to have a pattern of concern which will lead me to devote myself to helping the worst-off. The pattern that would lead me to do this would be quite impartially altruistic, i.e. lacking in the personal attachments which make one partial towards some people. But such a demand is much more than anything Common-sense Morality or the positive morality demands. In response to this objection, the consequentialist may appeal to Collective version of the theory. Collective Consequentialism tells me to have one of the patterns of concern which are such that, if everyone had one of these patterns, the best consequences would be produced. As I discussed in Chapter 2, such a pattern would not be so impartially altruistic. So Collective Consequentialism would not be so demanding as Act-consequentialism.

It is very difficult to determine how much a multi-level Collective Consequentialism would coincide with Common-sense Morality (and with the positive morality of our society). The blame for this should go to both sides. On the one side is the difficulty of deciding just what acts and dispositions are demanded by this sort of consequentialist theory. On the other

side is Common-sense Morality's imprecision.⁷ So perhaps I should put my conclusion in a conditional form: if some consequentialist theory does manage to coincide with Common-sense Morality, then obviously this consequentialist theory and Common-sense Morality match the positive morality of our society to an equal extent, in which case members of our society, with certain exceptions, would be just as prudent to accept this consequentialist theory as to accept Common-sense Morality. But if (as I suspect) no consequentialist theory matches our positive morality as well as Common-sense Morality does, then members of our society would, generally speaking, be more prudent to accept Common-sense Morality.

SECTION 4

I come now to the distinction between reciprocalist and unilateralist conceptions of morality. As I explained this distinction in Chapter 2, a reciprocalist conception of morality does not require us to treat others well unless they are, or at least would, treat us well; whereas, a unilateralist conception of morality requires us to treat others well even if they are not treating us well.

Of course we have merely to formulate the question 'Which is better from a prudential point of view, having unilateralist or having reciprocalist moral dispositions?' and the answer is immediately apparent—reciprocalist. I am undoubtedly much better off in self-interested terms if others treat me well (that is, if they do not hurt me, steal from me, coerce me, break their agreements with me, lie to me, or ignore me when I need

rescuing). But there are selfish people around who would not treat me well unless they were furnished with some self-interested reason for doing so. I can give them a self-interested reason for treating me well by making my treating them well conditional on their reciprocating. Since I need their good behaviour towards me and I make this behaviour more likely by being a reciprocalist rather than a unilateralist, I benefit from being a reciprocalist.⁸ (Of course, everyone else is like me in this respect.)

Let me mention here that all reciprocalist principles are agent-relative. Reciprocalist principles say that the agent is to act in certain ways towards others who act in these ways towards the agent. Within the reciprocalist principle a clause in which the agent is mentioned will be needed to specify those others. Consider, for example, the reciprocalist principle, 'morality requires the agent to be honest with others who are honest with the agent'. (Or, 'if and only if agent B is honest with agent A is A required to be honest with B.') Or consider a more general reciprocalist principle, 'morality requires the agent to treat well others who treat the agent well.'

Now does the point that reciprocalism is more prudent than unilateralism favour Common-sense Morality or consequentialism? Well, it follows from the definitions of consequentialism and agent-relativity that I gave in Chapter 2 that no form of consequentialism has first principles which are agent-relative. And, as I have just said, all reciprocalist principles are agent-relative. It follows that no form of

consequentialism is reciprocalist (at the level of first principles). Indeed, we can see that both utilitarianism and Distribution-sensitive Consequentialism are unilateralist: if your helping me would result in the greatest net welfare, or alternatively the greatest welfare and equality, then it would be wrong of you not to help me, and this holds independently of whether I would or will ever help you.

Some consequentialists may complain that I am wrong to say that their theory is unilateralist. These consequentialists might say that they would endorse our having the disposition not to treat people well who refuse to treat us well. They might endorse this disposition on the grounds that, if we had it, this would encourage others to treat us well, which would presumably bring about better consequences, impartially considered. But let us insist that, as well as saying what dispositions it is right to have, the consequentialist tell us which acts are right and wrong. The act-consequentialist believes that, when assessing the rightness or wrongness of an act, what matters is whether the act would produce the best consequences, not whether the beneficiary is disposed to follow the same principle. This is a unilateralist belief about the morality of the act. So, no matter what they say about dispositions, act-consequentialists are plainly unilateralists about the rightness and wrongness of acts.

Collective Consequentialism seems to be also essentially unilateralist in the relevant sense: the agent is required to do what, if everyone did it, would produce the best consequences. According to Collective Consequentialism, whether others are

actually following the principle that it would be best if everyone followed is irrelevant. So this theory would seem to hold that whether or not the prospective beneficiaries of one's act are actually disposed to reciprocate is irrelevant. (But it may not be altogether clear that Collective Consequentialism is unilateralist. What makes this difficult to decide is that Collective Consequentialism is difficult to formulate. This difficulty comes largely from the problem of knowing how specific Collective Consequentialism's principles are allowed to be. But I do not have room here to go into the complexities of this.)

Our question at the moment is, does the point that reciprocalism is more prudent than unilateralism favour Common-sense Morality or consequentialism? I have argued that consequentialism is unilateralist. Now, is Common-sense Morality unilateralist or reciprocalist? In Chapter 2, I cited Common-sense Morality as a deontological morality. I also said there that all deontological principles are agent-relative. I have explained earlier in this section that all reciprocalist principles are agent-relative. But, of course, it does not logically follow that all agent-relative principles are reciprocalist. Furthermore, we can think of an example of a principle that is agent-relative and yet unilateralist. Consider the Golden Rule, 'Do unto others as you would have them do unto you', which can be reformulated in order to bring out its agent-relativity as, 'If A would have B, C, D, E, etc. treat him or her in a certain way, then A is required to treat them in that way.' This principle is agent-relative because it must mention the agent in the statement of the conditions under which

the agent is required to do something (in the example, treat others in a certain way). It is unilateralist because it implies that, because you would have me be honest with you, you are required to be honest with me even if I am not actually honest with you.

So, then, assuming that the principles of Common-sense Morality are the principles of our positive morality, let us ask whether they are unilateralist? Well doesn't our positive morality include principles about loving thine enemies, turning the other cheek, and (more generally) treating others as you would have them treat you, and are not these principles basically unilateralist? But let us distinguish between, on the one side, a morality's ideals of supererogation and, on the other, its demands—the things people are expected to do and blamed or criticized for not doing. Having made this distinction, we might think that turning the other cheek, loving enemies, and (more generally) treating others as we would want to be treated, even when they are not treating us in these ways, are not what our positive morality demands, but rather are its supererogatory ideals.⁹ On the other hand, I do not know whether Common-sense Morality and our positive morality permits us (e.g.) to be dishonest in our dealings with people who are dishonest towards us. Earlier I said that Common-sense Morality is imprecise—the same holds for our positive morality. But, on the whole, it does seem that the principles of Common-sense Morality and our positive morality are at least largely reciprocalist. If that is correct, Common-sense Morality again comes out ahead of consequentialism (and ahead of any unilateralist deontological

principle, such as the Golden Rule).

SECTION 5

Presumably it would be prudent to accept a morality that not only gives one no duties towards those who choose not to reciprocate, but also gives one no duties towards those who are incapable of reciprocating. For how could it be to one's advantage to be disposed to act in the interests of (e.g.) the generations that will live in the far-off future? Or how could one profit from having dispositions to act for the benefit of animals?

Actually, there are conditions under which having such dispositions would be in one's interest—e.g. the moral code established in one's society might require such behaviour. Indeed, most people in our society believe that we have certain (however weak) obligations towards all living human beings even if they are and will remain too helpless to reciprocate. And they believe that it would be wrong to give no weight to the interests of future generations. Thus there are unilateralist elements in our positive morality. (And I take it that the same is true of Common-sense Morality.) Because of this fact about our positive morality, it would not be prudent for a member of our society to behave in a fully reciprocalist fashion.

Still, if the positive moral code of one's society contains no such requirements, prudence would seem to favour one's accepting a fully reciprocalist morality. But concluding that it is prudent to have a morality which gives one no duties to creatures who are incapable of reciprocating might open the door to important difficulties. For the rich and powerful might

say, 'there is no prudential advantage for us in making any serious sacrifices for the poor, who cannot comparably benefit us in return.' To this we can only reply that, if the poor really do have nothing comparable to offer the rich, then the rich may have no self-interested reason to accept a morality which obligates them to forgo advantages for the sake of the poor. But perhaps there is some reciprocity between the rich and the poor. G. Kavka suggests that if the rich treat the domestic poor badly, the poor might feel compelled to turn to robbery or revolution, which are at best expensive to stamp out; and that if the rich countries fail to give aid to the poor countries, there is the risk of 'military conflict started by some desperate poor nation, or some desperate group within such a nation'.¹⁰ Kavka concludes that, because of such reciprocity (in our age), even a fully reciprocalist conception of morality would require the rich to make some sacrifices for the poor.

However, one objection to Kavka's argument is that, even if it were in the interests of the rich as a class to accept a morality which required them to aid the poor, the individual rich person might be better off with a morality which didn't require him or her to contribute to this project. But let us set aside that point and turn to a second objection, namely, that Kavka's claims about the amount of reciprocity between the rich and poor may be too optimistic. To be sure, some rich do believe that it is in their best interests to 'buy off' the poor in their own country through charity and welfare programs. Others, however, seem to believe that, in the domestic sphere, it is not redistribution that is called for, but stiffer laws, more police,

and perhaps some new 'opiate for the masses'; and in the international sphere, better military and economic strategies 'to protect their vital interests'. If these people are correct, Kavka's claims about the amount of reciprocity between rich and poor are false. And if Kavka's claims are false, then (as I said) the rich may be better off with a morality that doesn't require them to aid the poor.

I have been emphasizing that the major factor in determining whether it is prudent to accept any one moral theory rather than its rivals is whether that moral theory is embodied in the socially established moral code. But there may be a problem here. My present beliefs might be such that, in order to come to accept the moral theory embodied in the positive morality, I would have to arrange for myself to be deceived in some way. This need for deception complicates matters. Most Objective List accounts of self-interest identify not being deceived as an intrinsic prudential good. Furthermore, Desire-fulfilment accounts of self-interest hold that, if one desires not to be deceived, then not being deceived is an prudential good. So, on these accounts, in order for the self-deception to be in my interest on the whole, the loss to me of being deceived would have to be less than the benefit to me of coming to accept the positive morality. But it seems that, no matter which of the plausible accounts of self-interest is correct, there will be some people for whom the prudential benefits of believing in the positive morality would be so large as to make it, on the whole, in their interest to undergo the deception.

SECTION 6

So far in this chapter the focus has been on the question: Which normative moral theory does prudence recommend that one accept? Normative moral theories are of course made up of beliefs about which acts and dispositions are right or wrong, admirable or blameworthy. But we can also ask what metaethical beliefs prudence recommends. (Again, we should note that someone who believes one metaethical view might be unable to come to believe some other metaethical view without arranging for some form of self-deception. But what I just said about the prudence of self-deception about normative beliefs will be applicable here as well. That is, self-deception about metaethics will at least sometimes be prudent.)

It seems that the metaethical view most of us would be prudent to accept is Conventionalism. I mean by 'Conventionalism' the thesis that the correctness of any normative moral belief is relative to the shared understandings and tacit conventions that actually obtain within some society.¹¹ I have already said that it is usually in one's interest to accept the positive morality of the society one lives in. Believing in Conventionalism is thus prudentially advantageous in that it helps one swallow whatever normative code happens to be embodied in this positive morality. The belief in Conventionalism could help us accept the positive morality of the society in which we grow up, and then if we migrate to a society whose positive morality is different, it could help us come to accept the new positive morality. On the other hand, some

positive moralities themselves evidently forbid the belief in Conventionalism, and even deploy external sanctions against those who fail to believe in some opposing metaethical view. (The example that springs to mind is of positive moralities which sent people to be burned at the stake for believing that what makes things right or wrong is social custom rather than divine command.) But, apart from such cases, the acceptance of Conventionalism seems likely to be in the agent's best interest.

SECTION 7

My main argument in this chapter was that, because Common-sense Morality both has a less demanding principle of aid than the main forms of consequentialism do and is less unilateralist than they are, and because Common-sense Morality rather than any form of consequentialism appears to be the positive morality of our society, a member of our society would be more prudent to accept Common-sense Morality than to accept consequentialism. Now if the correct or best justified moral beliefs are consequentialist ones, then from the point of view of self-interest it is not advisable to have the correct moral beliefs. Similarly, if some metaethical view other than Conventionalism is true, then from a self-interested perspective it is not advisable to hold the correct metaethical view. On the other hand, if Common-sense Morality is the correct normative theory and Conventionalism the correct metaethical theory, then, though (as I argued in the previous chapter) prudence would not recommend that one's moral desires be strong enough to override always self-love, prudence would at least recommend that one have the correct moral and

metaethical beliefs.

PART THREE

WHAT DO I HAVE MOST REASON TO DO?

CHAPTER 7

THE SELF-INTEREST THEORY VERSUS THE PRESENT-AIM THEORY

SECTION 1

In Part Two of this thesis I found that being moral may well not be in one's self-interest. But someone might say that even if being moral is not best for oneself, one might nevertheless have most reason to be moral. This claim faces us with the question 'What does one have most reason to do?' Part Three of the thesis explores this question.

I shall take this question to be equivalent to 'What are the strongest reasons for action?' Defining rational action (or, more accurately, supremely rational action) as action in accordance with the strongest applicable reasons for action, I sometimes formulate the question as 'What is it supremely rational to do?'

In the following section I outline some of the alternative possible theories of what one has most reason to do. As I shall soon be explaining, some of these theories deny that morality as such generates reasons for action. The present chapter will discuss certain arguments which might be thought to show that one particular theory which denies this is superior to the other theories which deny it. Then in the final chapter of the thesis, I will consider other arguments which are meant to show that this particular theory is superior to all other theories of practical reason, including the theory that morality does generate reasons for action.

Now, whether acting to promote X rather than Y is

supremely rational will often turn on the question of whether caring more about X than Y is supremely rational. To put the point in another way, theories of practical rationality disagree with one another not only about which acts one should do but also about which desires should be stronger than the others. It will facilitate the discussion of these competing theories if we have an expression which refers to a set of desires, aspirations, dispositions, etc., at certain relative intensities.

Following Parfit, I shall use the expression 'pattern of concern'.¹ So if you and I have exactly the same desires and dispositions, A, B, C, ..., but in your set the dominant desire is A while in my set it is B, then we have different patterns of concern.

SECTION 2

Let me now briefly introduce the main alternative theories of practical rationality. A claim some different complete theories would incorporate is that the requirements of the best justified morality provide the strongest reasons for action.

Note that this claim is about best justified morality.

In the previous part of this thesis I argued that it is primarily positive morality that is of direct relevance to the question of what is in one's self-interest. But in this part of the thesis the focus is on reasons for action. And I assume that it is the best justified morality that would be claimed to provide, on its own, reasons for action. So, from now on the only kind of morality I shall be concerned with will be the best justified (narrow) morality. To save words, I shall refer to it as

morality.

Why would the claim that moral requirements provide the strongest reasons for action not itself be a complete theory of reasons for action? A complete theory would be able to say, about any situation, what there is most reason to do. But, on many conceptions of morality, it is silent in some situations as to what the agent should do, thus being unable to give the agent any reason for action. However, the claim that moral requirements generate the strongest reasons can be combined with other claims to make up a complete theory. Someone might hold, for example, that when there is a moral requirement in play one has most reason to act morally and the rest of the time one has most reason to do what is best for oneself. Alternatively, it might be held that when there is a moral requirement in play one has most reason to act morally and the rest of the time one has most reason to do what will best fulfil one's present desires. But I want to postpone until the next chapter the question of whether moral requirements do generate the strongest reasons, or even any reasons at all.

One of the theories I will concentrate on in the present chapter is the Self-interest Theory, as defined by Parfit.² I will call this theory the Direct Self-interest Theory in order to distinguish it from a different theory that will be introduced shortly. The Direct Self-interest Theory insists that the supremely rational act is that which will make the agent's life, taken as an extended whole, go best for the agent. As I have mentioned, we can distinguish three general versions of the Direct Self-interest Theory. The most plausible of the

Hedonistic versions of this theory claims that the agent has most reason to do what will give him or her maximum enjoyment (i.e., the most pleasant succession of mental states possible). In contrast, Desire-fulfilment versions claim that the agent has most reason to do what will maximally fulfil the agent's present and future desires (whether or not this will bring the agent the most pleasant mental states). Lastly, Objective List versions of the Direct Self-interest Theory, maintain that the supremely rational act is that which brings the agent the things on the list, whether or not this act also brings the agent pleasant mental states or best fulfils the agent's desires.

A third theory is what I shall call the 'Indirect Self-interest Theory'. The best way to explain this theory is to contrast it with the Direct Self-interest Theory. The Direct Self-interest Theory holds not only that the supremely rational act is the act best in the long-run for the agent, but also that the supremely rational patterns of concern are those patterns within which self-love (the desire to do what is best for oneself) is dominant. But the Direct Self-interest Theory does not necessarily maintain that it is supremely rational for a person to develop or maintain a pattern within which self-love is dominant. (That is, the Direct Self-interest Theory does not necessarily maintain that it is rational to cause oneself to have the pattern of concern that the Direct Self-interest Theory itself claims is the supremely rational pattern of concern.) As I argued in Chapter 3, all plausible versions of the Direct Self-interest Theory can acknowledge that one might be better off in self-interested terms if some of one's non-self-interested

desires are so intense that they could overpower sometimes one's self-love and lead one to do acts one knows to be prudentially worse. But the acts that these patterns of concern would lead one to perform would not themselves be supremely rational according to the Direct Self-interest Theory. Any act is, on the Direct Self-interest Theory, irrational if it is not what is best for the agent—even if the act flows from a pattern of concern the having of which will give the agent the best possible life in self-interested terms. The Indirect Self-interest Theory, however, disagrees. It maintains that any act flowing from such a pattern is supremely rational—even if that particular act is not as good for the agent as some alternative would have been.³

But the three general alternatives mentioned so far—the Direct Self-interest Theory, the Indirect Self-interest Theory, and theories that incorporate moral reasons—do not exhaust the alternatives. There is also what Parfit calls the 'Present-aim Theory'.⁴ This theory holds that acts can be perfectly rational even when they are immoral or imprudent, and that patterns of concern can be perfectly rational even when they are not dominated either by moral desires or by the desire for what is best for oneself. So it conflicts both with the Direct Self-interest Theory and with theories holding that moral requirements generate the strongest reasons for action. (See the final note to my first chapter for complications.) The Present-aim theory also conflicts with the Indirect Self-interest Theory, for the former claims that an act can be supremely rational even if it is not one flowing from the

pattern of concern the having of which would make the agent's life go best in self-interested terms.

Just as there are different versions of the Direct Self-interest Theory and of the Indirect Self-interest Theory, there are different versions of the Present-aim Theory, as Parfit explains.⁵ The crudest version of the Present-aim Theory claims that one has most reason to do whatever will maximally satisfy one's strongest present desires overall, no matter what the present desires are. This version is concerned with the choice of the agent's means to his or her ends, but not at all with the ends themselves. However, a more plausible version of the Present-aim Theory—Parfit calls it the 'Deliberative' version—does criticize ends. It criticizes ends that the agent desires as a result of his or her mistaken empirical beliefs, faulty logic, or distorting experiences.⁶ The 'Critical' version of the Present-aim Theory goes further: it acknowledges that some desires are irrational even though they are not founded on false beliefs or faulty reasoning.⁷ Consider someone who wants a saucer of mud for itself rather than for any further reason.⁸ We might think that, even if this person's desire for a saucer of mud does not arise from distorting experiences or from mistakes in either reasoning or empirical beliefs, the desire still is irrational. Another example might be Rawls's case of someone who claims that he would rather count all the blades of grass in the lawns on the street than make great contributions to mathematics.⁹ Other examples of intrinsically irrational preferences come from cases in which someone draws an arbitrary line between possible objects of

concern.¹⁰ But the main point here is not whether such preferences are or are not genuinely irrational. Rather, the point is that the Critical version of the Present-aim Theory need not defend the rationality of such preferences.

Consider the contest between the Direct Self-interest Theory and the Deliberative Present-aim Theory. Many people assume that if one's reasoning and empirical beliefs were free from mistakes, one would most want to do what would promote one's long-term self-interest. This assumption might seem to support the view that there is no conflict in practice between the Deliberative Present-aim Theory and the Direct Self-interest Theory. We must be careful, however, to distinguish the following two ways of taking this assumption. On the first way of taking it, the assumption means that anyone who is thinking clearly will, as a matter of empirical fact, be most concerned to do what is best for himself or herself. This is the doctrine of Psychological Egoism. As I stated in Chapter 3, introspection and experience indicate that Psychological Egoism is empirically false. And Psychological Egoism, being false, cannot really provide any support for the claim that if one's logic and empirical beliefs were free from mistakes, one would most want what is best in the long run for oneself—cannot really support, that is, the claim that there is no conflict between the Direct Self-interest Theory and the Deliberative Present-aim Theory. So let us move to the second way of taking the assumption in question. According to this second way, most people have a standing desire that their own long-term future go well. But we can readily accept this without having to deny that the Direct

Self-interest Theory and the Present-aim Theory conflict in practice, as well as in theory. Though most of us are strongly motivated to do what is best for ourselves, this is not always our strongest present desire. And when it isn't, the Direct Self-interest Theory and the Deliberative Present-aim Theory are in practical conflict.

In the face of these claims, advocates of the Direct Self-interest Theory might shift their attentions to the Critical version of the Present-aim Theory, claiming first that it is superior to its cousin the Deliberative version, and then that there is no conflict in practice between the Critical version and their own Direct Self-interest Theory. They would appeal to the idea that the Critical Present-aim Theory acknowledges that some preferences would be irrational even if they were not based on faulty reasoning or mistakes about the empirical facts. 'To care more about something else than one does about self-interest', those sympathetic to the Direct Self-interest Theory would then say, 'is just to have one such irrational preference'. Meanwhile, advocates of the Indirect Self-interest Theory would say, 'To care more about A than B though it is in one's self-interest to care more about B than A is to have an irrational preference.' But I shall assume that advocates of the Critical Present-aim Theory will resist these claims. Not to resist them would be to let the Critical Present-aim Theory collapse into either the Direct or the Indirect Self-interest Theory. If the Critical Present-aim Theory were to collapse in this way, it would lose interest.

There is, however, no harm in letting the Critical

Present-aim Theory accept that it is irrational to care more about one's own short-term good than one's own long-term good. The Critical Present-aim Theory can accept this much from the Direct Self-interest Theory.¹¹ What the the Critical Present-aim Theory, if it is to avert a collapse, must disagree with the Direct Self-interest Theory about is the rationality of caring more about other things (not including short-term good) than about one's own long-term good. While the Direct Self-interest Theory insists that such a pattern of concern would be irrational, the Critical Present-aim Theory maintains that it would not.

SECTION 3

Parfit advances three arguments in favour of preferring the Present-aim Theory to the Direct Self-interest Theory. One is a formal argument; another is that we have certain beliefs which are inconsistent with the the Direct Self-interest Theory; the third is based on his views about personal identity. In this thesis I do not have room to consider controversies about personal identity and their implications for practical rationality. But in the next section I consider Parfit's argument that we have certain beliefs inconsistent with the Direct Self-interest Theory. And in the present section I examine Parfit's formal argument against the Direct Self-interest Theory.

The formal argument starts from an observation about the formal differences in practical reasons. As I explained in Chapter 2, we can distinguish between moral principles that are

agent-relative in form and moral principles that are agent-neutral in form. To put the distinction briefly: an agent-relative principle involves an essential reference to the agent in the statement of the conditions under which the agent is required to promote a certain kind of event or outcome; this is not true of agent-neutral principles. Assuming for the moment that moral principles generate reasons for action, we can see that the formal character of the moral principles will carry over into the formal character of the reasons for action they generate. Consider, for example, the agent-relative principle: For all agents A and events E, if there is a promise which was made by A and E is the event which would fulfil that promise, then A is morally required to promote E. (As before, the reference to the agent essential for agent-relativity is underlined.) The corresponding agent-relative reason would be: that A's own promises be fulfilled (or more elaborately: For all A and E, if there is a promise which was made by A and E is the event which would fulfil that promise, then A has a reason, or there is a reason for A, to promote E.) This reason, coming from an agent-relative moral principle, is itself agent-relative because its full specification contains an essential reference to the agent in the statement of the conditions under which the agent has the reason. In contrast, a reason is agent-neutral if there is no mention of the agent in the statement of the conditions under which the agent has that reason.

Just as principles about what is morally required might apply universally, so might both agent-relative and agent-neutral

reasons apply universally. If they do, then we all have the same reasons. But an important point is that agent-neutral reasons for promoting outcomes are reasons for everyone to promote those outcomes; whereas one person might have an agent-relative reason for promoting an outcome which other people have no agent-relative reason for promoting.¹² Consider the outcome of my grandparents' having enough heat in their house. And consider a possible agent-neutral reason: that elderly people have enough heat. If this reason exists, then everyone has a reason for ensuring that my grandparents' have enough heat in their house today. Now suppose instead that there is only the agent-relative reason: that one's own grandparents have enough heat. Under this assumption, while I have a reason for making sure my grandparents have enough heat in their house, you do not have a reason for doing this.

Reasons might not only be relative to agents but also to times. If the reason I have for promoting something is agent-relative, others might not have any reason for doing this. Analogously, if the reason I have now for promoting something is time-relative, I may not have at other times any reason for doing this.¹³

Now Parfit's formal argument starts from the following observations: the Direct Self-interest Theory points to reasons that are agent-relative but time-neutral; consequentialism points to reasons that are completely neutral, i.e. both agent- and time-neutral; and Common-sense Morality and the Present-aim Theory point to reasons that are completely relative, i.e. both agent- and time-relative.¹⁴

He then suggests that if the Direct Self-interest Theory is defended against consequentialism by the claim that reasons can be relative to agents, the Common-sense Moralist and the Present-aim Theorist will challenge the Direct Self-interest Theory with the claim that reasons can be completely relative, i.e., relative to times as well as to agents. On the other hand, if the Direct Self-interest Theory is defended against these last two opponents by the claim that reasons must be time-neutral, the consequentialist will add that reasons should be completely neutral, i.e., neutral with respect to agents as well as with respect to times.¹⁵ Parfit concludes, 'Either reasons for acting can be relative, or they cannot.... if they can be agent-relative, they can be relative to the agent at the time of acting.'¹⁶

He is right, of course, that the Direct Self-interest Theory is neither completely relative nor completely neutral. But, supporters of the Direct Self-interest Theory may think that this is no objection to their theory. They may say that their theory is correct precisely because it is neutral where it should be neutral and relative where it should be relative. They might say, 'Because you and I are different people, you have reasons to try to achieve your aims while I have reasons to try to achieve my aims. That is where consequentialism goes wrong—in not giving differences in personal identity a significance they have. And because I am now the same person I shall be in the future, it is supremely rational for me to give my future aims equal weight with my present aims. That is where the Present-aim Theory and Common-Sense Morality go wrong—in giving to

differences in time a significance they lack.' Thus, the Direct Self-interest Theory might be defended on the grounds that only it reflects the fact that the relation between me now and me later is utterly different from the relation between me and someone else.

According to Parfit, the Self-interest Theorist's reply that the two relations are utterly different is mistaken. Parfit claims, 'These relations are, in most respects, radically different. . . . But these relations may still be similar in the relevant respects'.¹⁷ Parfit then suggests not only that these relations must be treated similarly, but also that the treatment they both should receive indicates that the Present-aim Theory is the best theory. This argument goes as follows. Consider certain facts about what is rational in the way of belief, that is, what is rational in the sphere of theoretical rationality. The relation between different people is relevant to the question of whether it is rational for Henry to believe that p. Suppose that James believes that not-p. If Henry is the same person as James, it is irrational for Henry to believe that p. If Henry is not the same person as James, it can be rational for Henry to believe that p. Thus there is a sense in which the rationality of having a certain belief is agent-relative. When a person believes that not-p, it is irrational for this person to believe that p, and yet it can be rational for a different person (who does not believe that not-p) to believe that p. Moreover, theoretical rationality is also time-relative. Suppose that I now believe that p. Then, though it would be irrational for me to

believe now that not-p, it would not be irrational for me later to believe that not-p (when I no longer believe that p). Thus both the relation between different people and the relation between a single person at different times are crucially relevant to questions about what beliefs are rational. Parfit's suggestion is that just as both of these relations are crucially relevant to questions about which beliefs are rational, both are crucially relevant to questions about which desires and actions are rational.¹⁸

I shall now offer two objections to this argument. The first objection grants that theoretical rationality is not only agent- but also time-relative but maintains nevertheless that this is not true of practical rationality. Even if Parfit is correct that the relation between different people and the relation between one person at different times are both relevant to questions about what beliefs are rational, this does not prove that both relations are relevant to questions about what desires and actions are rational. All the defender of the Direct Self-interest Theory has to do is to argue that the case of practical rationality is not analogous to the case of theoretical rationality. The idea that there are such disanalogies is not new. To take one example, some have thought that while it is irrational to accept empirical beliefs one knows to be incompatible, it is not irrational to accept propositions about what ought to be done that one knows to be incompatible.¹⁹

The possibility that there might be a relevant disanalogy provides us with an objection to Parfit's formal argument against

the Direct Self-interest Theory. But I admit that this objection is not compelling. It would be compelling only if I could put my finger on a relevant disanalogy. This I have not done.²⁰

Instead, I shall raise a second and more telling objection to Parfit's formal argument. My second objection is not vulnerable to the retort that I haven't cited a relevant disanalogy, for this second objection itself presumes that there is no relevant disanalogy.

It will help if, before I state the second objection, I add some remarks about how Parfit's formal argument operates. As I have indicated. Parfit appeals to facts about theoretical rationality in order to support, in the sphere of practical rationality, the Present-aim Theory over the Self-interest Theory. We might say that this support is supposed to be forthcoming in two ways. The first, which Parfit explicitly mentions, is that what he says about theoretical rationality shows that, in the area of practical rationality, differences in timing should be given the same treatment as differences in personal identity. This seems to support the Present-aim Theory over the Self-interest Theory, because, as I explained, the Present-aim Theory does treat these two kinds of difference in the same way, while the Self-interest Theory does not. The second and more important way, which Parfit does not explicitly mention, is that what he says about theoretical rationality suggests that they should both get treated as fundamentally significant. That is, the treatment both differences should get renders rationality agent- and time-relative.

Well, is it true that both differences are fundamentally

significant, such that theoretical rationality is both agent- and time-relative? Certainly it seems to be true that I might have (a set of consistent) beliefs which are inconsistent with the different (set of consistent) beliefs that you hold. And, I might now have (a set of consistent) beliefs which are inconsistent with the different (set of consistent) beliefs that I shall hold in the future. Furthermore, it is certainly true that many writers think of the claims of rationality as being nothing more than the claims of consistency. If rationality is thought of in this way, then both theoretical rationality and practical rationality are indeed agent- and time-relative.

But consider the distinction between objective and subjective rationality. Objective rationality concerns what is rational given true beliefs. Subjective rationality concerns what is rational given what the agent believes (and perhaps available information). Parfit himself distinguishes 'between what we have most reason to do, and what it would be rational for us to do, given what we believe, or ought to believe.'²¹ And he himself sees that the dispute between the Direct Self-interest Theory and the Present-aim Theory is about what one really has most reason to do, not about what it is rational in the light of one's beliefs to do.

Now, keeping in mind that we are interested in the Present-aim Theory, the Self-interest Theory, and so on, as theories about objective practical rationality, let us look again at the facts about theoretical rationality that Parfit appeals to. Parfit considers what it is rational for a person to believe given what the person already believes at the time.²² That

is, the sort of theoretical rationality he is thinking about is subjective theoretical rationality. But the Present-aim Theory and the Self-interest Theory are, as I said, theories of objective practical rationality. Therefore, appealing to facts about subjective theoretical rationality, as Parfit does, is less appropriate than appealing to facts about objective theoretical rationality would be. So let us now ask: could it be objectively rational either for different people or for one person at different times to have inconsistent beliefs? Since objective rationality concerns what is rational given true beliefs, the objectively rational beliefs will just be whatever beliefs are true. And true beliefs about a particular event at a particular time remain true independently of either whom they are believed by or when they are believed. Objective theoretical rationality, in contrast with subjective theoretical rationality, is thus both agent- and time-neutral.

Where does this leave us with respect to the question of which is the best theory of practical rationality? I said that Parfit appealed to facts about theoretical rationality to support the idea that, in the sphere of practical rationality, differences in identity and differences in timing should both be treated as having fundamental significance, which is what the Present-aim Theory but not the Self-interest Theory claims. But given that he wants to discover truths about objective practical rationality by looking at theoretical rationality, he should have looked at objective theoretical rationality, rather than subjective theoretical rationality. From considering objective theoretical rationality, what we would infer about objective

practical rationality is that neither differences in personal identity nor differences in time play a crucial part in determining what reasons one has. Thus, we would infer that practical rationality is time- and agent-neutral. But this directly contradicts Present-aim Theory, which makes practical rationality out to be time- and agent-relative. Thus Parfit's formal argument seems to undermine rather than support the Present-aim Theory.

SECTION 4

Having rejected Parfit's formal argument for the Present-aim Theory over the Direct Self-interest Theory, I now shall take up his argument that we have certain beliefs incompatible with acceptance of the Direct Self-interest Theory.

A note of caution: in what follows I shall assume morality is silent. In order to decide which is superior, the Self-interest Theory or the Present-aim Theory, we should consider test cases in which morality does not lend its support to one of the two alternative theories.²³ If we were only to consider cases about which morality and the Direct Self-interest Theory happen to agree against the Present-aim Theory, we would not be able to evaluate the Direct Self-interest Theory on its own merits. On the other side, it would not be fair to the Direct Self-interest Theory to consider cases about which morality happens to agree with the Present-aim Theory.

Well, apart from cases where morality teams up with the Present-aim Theory, are there any cases about which we side with the Present-aim Theory against the Direct Self-interest Theory?

Parfit directs our attention to a case where someone cares more about a particular achievement (say, writing a brilliant novel) than about self-interest.²⁴ (Note that the issue is whether it is rational to care more strongly about some achievement than about having the life that is best from the point of view of self-interest. Remember that the Direct Self-interest Theory can allow that a rational set of desires will contain desires other than self-love. What the Direct Self-interest Theory holds to be irrational are not other desires per se, but other desires at an intensity strong enough to enable them to prevail over one's self-love.²⁵) So our question is: do we believe that it is rational to care more about some particular achievement than about self-interest?

To answer this question we need to take one version of the Direct Self-interest Theory at a time. Let us start with Hedonism. This theory admits that we have desires for things other than our own enjoyment, or pleasant mental states. The Hedonistic version of the Direct Self-interest Theory merely claims that it is irrational to care more about these other things than one does about the quality of one's own mental states. Yet many people believe that such preferences are not irrational. This belief would lead them to abandon the Hedonistic version of the Direct Self-interest Theory.

But rejecting the belief that it is irrational to care more about other things than one does about one's mental states does not entail the rejection of all versions of the Direct Self-interest Theory. For example, the Desire-fulfilment version of the Direct Self-interest Theory does not claim that it is

irrational to prefer something even if this would give one less good mental states. But let us now ask if this version accords with our considered beliefs about what preferences are rational. We should remember that the Desire-fulfilment version of the Direct Self-interest Theory claims that a pattern of concern is irrational if it is not dominated by the desire to do what best fulfils one's own present and future preferences considered impartially. But just as we might believe that it is not irrational to care more about (e.g.) certain achievements than about having pleasant mental states for oneself, so we might believe that it is not irrational to care more about those achievements than one does about what best fulfils one's own present and future desires considered impartially. And if this is what we believe, we must reject not only Hedonistic versions of the Direct Self-interest Theory but also Desire-fulfilment versions.

Parfit suggests that this same belief might seem to lead us not just to new versions of the Direct Self-interest Theory but to a completely different theory, namely, the (Critical) Present-aim Theory. But is the Present-aim Theory the only theory that can accommodate our beliefs about the desire for achievement? In this section and the next I shall discuss three suggestions that might make us less confident that the move to the Present-aim Theory is necessary.

The first involves the Indirect Self-interest Theory. As we have seen, the Indirect Self-interest Theory agrees with its cousin the Direct Self-interest Theory that it is rational to cause oneself to have whatever pattern of concern the having of

which will make one's life go best, even if the pattern contains some desires that are strong enough to outpull self-love.

However, the Direct Self-interest Theory insists that, no matter what patterns it is rational to cause oneself to have, not doing what is best for oneself is irrational, as is not caring most about what is best for oneself. The Indirect Self-interest Theory dissents from this by claiming that any act which issues from a pattern of concern the having of which will make one's life go best, on the whole, is a rational act, even if the act itself is not in one's self-interest.

We are discussing our conviction that an act can be rational even if it will not maximize the agent's pleasure or desire-fulfilment. Obviously, this conviction conflicts with Hedonistic and Desire-fulfilment versions of the Direct Self-interest Theory. On the other hand, of our beliefs about the rationality of various acts and patterns of concern, those in which we have the most confidence might be completely consistent with Hedonistic or Desire-fulfilment versions of the Indirect Self-interest Theory. If this is true, the convictions that push us to reject these versions of the Direct Self-interest Theory might not push us to reject these versions of the Indirect Self-interest Theory.

Consider the following example. John, a novelist, cares more about writing a great novel than he cares about doing what will give him the life that is best in either hedonistic, desire-fulfilment, or objective list terms. If John acts on this pattern of concern, his act is irrational according to the Direct Self-interest Theory. Let us now add that if John caused himself

to have a different pattern of concern, his life would go better in self-interested terms. So the Indirect Self-interest Theory would say that John's acting on his present, non-optimific pattern, is irrational. Does the Indirect Self-interest Theory cohere with our beliefs about this case? When we learn that John could have decided to develop in himself a pattern more beneficial to himself, perhaps we are less certain that his seeking achievement at the cost of self-interest is rational. If we are less certain, it seems we are drifting into the camp of the Indirect Self-interest Theorist.

But this camp is not one we can stay in for long. For there are conclusive arguments for rejecting the Indirect Self-interest Theory. Since others present these arguments so convincingly,²⁶ I shall merely outline them. The Indirect Self-interest Theory maintains that any act one would be led to perform by a pattern of concern which one was prudent to cause oneself to have would thereby be itself a rational act. All we need to do to refute this theory is to describe an act that is clearly irrational and then tell a story where someone is led to do this act by a pattern of concern he or she was prudent to cause himself or herself to have.

Schelling discusses the making of threats, where threats are defined as promises to carry out mutually harmful acts in the event that someone else behaves in a certain way. The fulfilment of these threats harms everyone, including the agent. Let us modify the standard example somewhat by stipulating that the agent never would have occasion to interact either with anyone who has ignored one of the agent's previous threats or with

anyone who has known of such an occurrence. This stipulation is made in order to rule out any possibility of an advantage accruing to the agent from an enhanced reputation as a fulfiller of threats. Given these stipulations, those who threaten to do such acts would have no self-interested reason to carry them out. Although carrying out such a threat would be highly imprudent, it can still be prudent to make such a threat because, in Schelling's words, 'the threat and not its fulfilment gains the end.'²⁷ It gains the end, makes one better off, by inciting others to make sure that the circumstances in which one would carry out the threat never (or at least, rarely) obtain.

But because I would have no incentive to carry out the threat in the unhappy event that my threat is ignored, no one would believe my threat unless I could provide some evidence that I would carry it out. Given this fact, and given that I could not make myself into a convincing liar, it might well be in my best interests over the long-term to cause myself to have an irresistible compulsion to carry out my threats. Now this compulsion would lead me in certain circumstances to do acts that would harm everyone concerned (including me) with no benefit to anyone. On the particular occasions when I act on my disposition to fulfil my threats, thereby harming everyone including myself, my act is clearly irrational. Thus, an act can be irrational although causing oneself to be disposed to perform acts of that kind was prudent. We cannot, then, accept the Indirect Self-interest Theory.

SECTION 5

I shall now outline a different defense of the Direct Self-interest Theory. In order to explain this defense I need first to mention that (as Parfit explains) the Direct Self-interest Theory might tell me to try to acquire certain beliefs which must be false if the Direct Self-interest Theory itself is true.²⁸ Suppose that it is in my interests to have a strong desire for achievement, so strong that it would sometimes get its way against my desire to do what is best for myself. As long as I believe that no act is rational unless it is best for the agent in self-interested terms, I might be psychologically incapable of desiring achievement with the intensity necessary for me to have the most rewarding life possible. Thus the Direct Self-interest Theory might tell me to cause myself to believe that any act is rational either if it is best for the agent or if it best fulfils some desire for achievement. But, of course, this belief is false if the Direct Self-interest Theory is true. (I shall not repeat here what, in the preceding chapter, I said about the complications introduced by self-deception.) So, if I can't simultaneously hold obviously inconsistent beliefs, the Direct Self-interest Theory tells me to cause myself to believe some other theory.

I can now explain the defense of the Direct Self-interest Theory. The defense consists of explaining away our belief that acting on a desire for achievement is rational even when this will result in a net loss to self-interest. The defender of the Direct Self-interest Theory might say that we have this

belief because it is good for us, in self-interested terms, to have it. The argument employed here by the defender of the Direct Self-interest Theory is like Hare's argument that there are good utilitarian reasons for our having the counter-utilitarian intuitions we do in fact have.²⁹ The defender of utilitarianism might admit that we have a conviction that hanging an innocent person is wrong even when this act would produce the greatest net happiness impartially considered, and yet argue that this conviction is no weapon against utilitarianism because utilitarianism itself endorses our having it. Likewise, the defender of the Direct Self-interest Theory might admit we have a conviction that trying for some achievement is not irrational even when this act is not what is best for the agent in self-interested terms, and yet argue that this conviction is no weapon against the Direct Self-interest Theory because this theory itself endorses our having it.

But an objection to this defense might be that we have never in fact been believers in the Direct Self-interest Theory. And how could the Direct Self-interest Theory play a causal role in our coming to have these beliefs if we have never believed in that theory? There may be an answer to this. Perhaps our ancestors accepted the Direct Self-interest Theory; then, on the Direct Self-interest Theory's instructions, acquired these beliefs that are inconsistent with the Direct Self-interest Theory; then found that they could not maintain these beliefs while simultaneously believing in the Direct Self-interest Theory; and therefore caused themselves to lose their belief in the Direct Self-interest Theory; and then, finally, taught their

descendants (us) that the Direct Self-interest Theory was false and that the other beliefs were true. An alternative historical account might be that they themselves never stopped believing in the Direct Self-interest Theory, but nevertheless concealed this from us in order that our commitment to the other beliefs, which they did pass on to us, could be stronger.

The first alternative historical account is unconvincing. What is wrong with it is not that our ancestors were not believers in the Direct Self-interest Theory—indeed belief in the Direct Self-interest Theory might well have been prevalent for a very long time.³⁰ Rather, the problem is that the first account seems to attribute to our ancestors a process of self-deception it is difficult to believe many of them went through. The suggestion that, on a widespread scale, our ancestors intentionally deceived themselves into having certain beliefs inconsistent with the Direct Self-interest Theory is at least slightly ridiculous. The second alternative historical account described above, though less unbelievable, is extremely speculative and cries out for historical evidence. Therefore, this general line of defense of the Direct Self-interest Theory is unpersuasive, though interesting.

SECTION 6

I have now discussed two ways that, in the face of Parfit's argument that we have beliefs which conflict with the Direct Self-interest Theory, we might resist moving to the Present-aim Theory. The first involved moving to the Indirect Self-interest Theory. But I argued that this theory is, ultimately,

unacceptable. The second option involved defending the Direct Self-interest Theory on the grounds that it can itself explain away beliefs that we have which seem to conflict with it. I concluded that this defense is not altogether satisfying. In this section I shall explore a third way that, without moving to the Present-aim Theory, we might accommodate the belief that it can be rational to act on one's desire for some achievement even when this act is not best for oneself in either hedonistic or desire-fulfilment terms.

It consists of maintaining that what we really believe is that only when one's present desires are for certain ends does one have more reason for acting on these desires than for doing what would make one's life go best in either hedonistic or desire-fulfilment terms. That is, we might agree with Parfit that it is not irrational to care more about achieving something truly significant than about maximizing one's pleasure or the fulfilment of one's present and future desires considered impartially. When we think about pleasure as such, independent of what produces this pleasure, we see that it is not irrational to care less about pleasure than about other ends, such as an achievement. Similarly, when we think of desire-fulfilment as such, i.e. ignoring the content of the desires, we see that it is not irrational to care less about it than about certain other things, such as certain achievements. Finding that it is not irrational to care more about certain things such as achievement does not, however, force us to accept that the Present-aim Theory is superior to the Direct Self-interest Theory. When we ask ourselves what exactly are

these ends which it is not irrational to care more about than pleasure and desire-fulfilment, we find ourselves cataloguing just the goods that would appear on a plausible Objective List Theory of self-interest, goods such as achievement, friendship, knowledge, and perhaps the appreciation of beauty. That is, we might think that it is rational for one to do what will give one less good mental states and provide for less fulfilment of one's present and future desires only if this will bring one great achievement or friendship or knowledge. If we do believe this, we are forced to abandon not all versions of the Direct Self-interest Theory, but only the Hedonistic and Desire-fulfilment versions--our beliefs might be perfectly consistent with the Objective List version of the Direct Self-interest Theory. Of course, if we do accept the Objective List version of the Direct Self-interest Theory, we must maintain that it is irrational to do what one knows to be worse, in terms of the things on the list, for oneself. But this belief is not inconsistent with the beliefs mentioned above.

What we could not maintain, should we accept this theory, is that it is rational to care more about, say, achievement than about one's own overall good. Of course, a plausible Objective List account of self-interest will list a plurality of goods, e.g., achievement, friendship, knowledge, and perhaps the appreciation of beauty. This may give rise to an objection to the Objective List version of the Direct Self-interest Theory. Parfit writes,

It may be true that, to fulfil my desire for achievement, I must deny myself most of the other things that are good for me. I may thus be doing what, on the Objective List

Theory, will be worse for me. But, since I am fulfilling a desire for achievement, ... I am not acting irrationally.³¹

The conclusion Parfit draws from this is that we must reject the Objective List version of the Direct Self-interest Theory in favour of the Present-aim Theory.

In reply, I acknowledge that, if the Objective List version of the Direct Self-interest Theory implies what Parfit says it does, then he is correct that the theory seems wrong. But let us distinguish between Specializing and All-rounder conceptions of the Objective List Theory. All-rounder conceptions would hold that what is in one's self-interest is having roughly equal amounts of each thing on the list (given that some threshold amount of ~~certain~~ things have been attained). It is this sort of conception that is particularly vulnerable to Parfit's objection. But now consider the Specializing conception, according to which (e.g.) the most magnificent achievements compensate the agent for some deficiencies in the other goods. So conceived, the Objective List Theory is much less vulnerable.

Of course, there may be levels of deficiencies so severe that, even according to the Specializing conception, nothing could compensate for them. Thus the Specializing conception will claim that it is always irrational to act in a way that brings about such deficiencies, no matter what the amount of other goods the agent obtains. We may be unwilling to accept this claim, however, because we hold that when there is a moral reason for doing a particular act, doing it can be supremely rational even if there will be no compensation to self-interest. What may make

it rational to pursue great achievement at the expense of the other things on the list is the benefit others get from the great achievement. But if this is our thinking, we are rejecting the Specializing Objective List version of the Direct Self-interest Theory not in favour of some Present-aim Theory, but in favour of some theory which insists that morality can yield reasons for action stronger than those yielded by self-interest. However, I want hold off discussing moral reasons until the next chapter.

What would show the Specializing Objective List version of the Direct Self-interest Theory to be inferior to the Present-aim Theory is an example in which (a) the agent clearly would make his or her life go worse by fulfilling his or her present desire for some one particular good on the list rather than doing what will bring him or her the rest of the goods on the list, and (b) the agent nevertheless has at least as much reason to seek the one good as he or she has to seek instead the rest of the goods, and (c) this reason is not a moral one. I cannot think of such an example.

SECTION 7

To sum up the present chapter: I have assessed two of Parfit's arguments for concluding that the Present-aim Theory is superior to the Self-interest Theory. His formal argument was subjected to two objections. Those objections showed, I think, that his formal argument does not in fact prove that the Present-aim Theory is superior to the Self-interest Theory. The other argument of his that I addressed was that we believe it is not irrational to care more about certain things (e.g., achievement)

than about self-interest, and that this belief is incompatible with the Self-interest Theory. I suggested that what we actually believe is instead that it is not irrational to care more about these things than about either pleasure or the maximum fulfilment of one's present and future desires considered impartially. I said that the Direct Self-interest Theorist (of any variety) might try to explain away this belief, but I doubted whether such an attempt would be convincing. Much more promising is the idea that the belief could be accommodated by Objective List versions of the Direct Self-interest Theory, especially by Specializing conceptions. (As I warned, I did not consider Parfit's argument from personal identity.) So, setting aside his argument from personal identity, we should conclude that Parfit has not shown the Present-aim Theory to be superior to the Self-interest Theory.

CHAPTER 8

DO I HAVE MOST REASON TO BE MORAL?

SECTION 1

In the previous chapter I attacked two arguments for favouring the Present-aim Theory over the Self-interest Theory. But are there other arguments which show that the Present-aim Theory is superior to all other theories of practical rationality—i.e. superior not only to the Self-interest Theory but also (and more importantly) to theories which say that morality generates reasons? To put the question another way: Does morality generate reasons for action? But even if we are persuaded that it does, and thus that there is always some reason for acting morally, there is then the further (difficult) question of whether there is always more reason for acting morally than for doing anything else. In this chapter, I shall try to answer these questions.

I start with some preliminary remarks. First, as I indicated in Chapter 2, I shall concentrate on moral requirements, i.e., ordinary duties, and ignore for the most part (what might be called) ideals of supererogation, which are more demanding than morality's ordinary requirements. The claim that morality generates the strongest reasons for action is much more plausible if it means merely that moral requirements generate the strongest reasons than if it means that all moral considerations, including ideals of supererogation, generate the strongest reasons. Thus, I focus on moral requirements.

Second, let me acknowledge that both the Direct Self-interest Theory and the Present-aim Theory allow that on

some occasions we do have reason to do what we are morally required to do. (I don't mention the Indirect Self-interest Theory in this chapter since it was decisively refuted in the last.) The Direct Self-interest Theory claims we on some occasions have reason to do an act morality requires, namely, when the act would serve our self-interest; but this reason for doing what morality requires derives from self-interest, not from morality. Similarly, the Present-aim Theory allows that we on some occasions have reason to do what morality requires, namely, when it would best fulfil our present desires; but here the reason derives from the present desires, not from morality. Thus, the Direct Self-interest Theory and the Present-aim Theory (as I have defined them) deny that morality itself generates reasons.

Finally, as I indicated in the last chapter, morality, according to many conceptions of it, is sometimes silent. In such cases, the agent needs some nonmoral reason for choosing one alternative over the other. Presumably, either the agent's present aims or long-term welfare provide the needed reasons in these cases. Hence, if it is correct that morality is sometimes silent, moral requirements are not the only things that generate reasons for action. And even if morality is never silent, there may nevertheless be other sources of reasons. So I shall take it for granted that morality is at most one of the sources.

SECTION 2

In this section, I outline four alternative conceptions, found in the literature, of the relations between moral requirements, reasons for action, and motivation. The first claims that one

cannot be morally required to do something one has no reason for doing, and that one cannot have a reason for doing something unless one has (independently) a motive for doing it. Or:

- (1) Having an existing desire that would be fulfilled if one does X is a necessary condition for having a reason for doing X, and having a reason for doing X is a necessary condition for being morally required to do X.

Since 'is a necessary condition of' is a transitive relation, having a present desire that would be fulfilled by the act is, according to this view, a necessary condition of being morally required to do the act. View (1) is essentially the Present-aim Theory about reasons plus the claim that moral requirements too are determined by present desires. Harman is an advocate of this view.¹

The second view to be considered is:

- (2) Having an existing desire that would be fulfilled if one does X is a necessary condition for having a reason for doing X, but having a reason for doing X is not a necessary condition for being morally required to do X.

That is, (2) agrees with (1) about the relation between reasons for action and desires but not about the relation between reasons for action and moral requirements. To be more specific, (2) goes along with (1) in accepting the Present-aim Theory about reasons for action. But (2) does not go along with (1)'s claim that we cannot be morally required to do some act unless we have a reason (as the Present-aim Theory conceives of 'having a reason') for doing that act. (2) insists that a moral requirement can obtain even in the absence of a reason for action.²

Now that I have listed the two views, of those I shall be discussing, that accept the Present-aim Theory, let me mention an

important complication in that theory. Advocates of the Present-aim Theory acknowledge that when one discovers new empirical facts or reflects on one's antecedent motivations, a new set of desires can arise, which can then in turn generate new reasons for action.³ This is an important acknowledgement, for, while it makes the Present-aim Theory more plausible, it also threatens to blur the contrast between the Present-aim Theory and some of its rivals. I shall come back to this problem later. But let me say here that in order to save words I shall use the term 'present desires' to refer to the set of antecedent desires complemented by new desires that would come either from new empirical beliefs or from reflection on one's antecedent desires.

A third view is that being morally required to do a particular act is a sufficient condition of one's having a reason for doing that act, and that one's having a reason for doing a particular act is a sufficient condition for one's having, if one is rational, a motive for doing that act. In other words:

- (3) Moral requirements always generate reasons for action, and these reasons supply the (objectively) rational agent with motivations.

This view seems to be held by some moral realists.⁴

(1) talks of desires as necessary conditions of both reasons and moral requirements, and (3) talks of requirements as sufficient conditions of both reasons and motivations. But since 'X is a necessary condition of Y' implies 'Y is a sufficient condition of X', it may seem that (1) and (3) don't really differ. However, they do differ: (1) refers to (what I said above I would be calling) present desires, whereas (3) refers to

motivations an objectively rational agent would have.

(2) goes along with (1)'s claim that having an existing desire that would be fulfilled if one does X is a necessary condition of having a reason for doing X. I shall discuss in the following section an argument that leads people to accept this claim.

The fourth and final view I shall mention is:

- (4) Moral requirements generate reasons for action, but failure to be moved by reasons isn't irrational.

D.A.J. Richards seems to be an adherent of this view.⁵ Note that (4), like (3), accepts that moral requirements generate reasons. (4) disagrees with (3) only about the relation between reasons and rational motivation.

SECTION 3

(1), (2), and (3) all accept that something can be a reason for action only if it is capable of motivating.⁶ But (1) and (2) accept the further claim that reasons for action are tied to the agent's present desires. (As I said, by 'the agent's present desires' I here mean the desires the agent would have after any mistakes in his or her logical reasoning and empirical beliefs have been corrected.) (3) does not accept this further claim. Why do the two sides disagree over this further claim? Followers of (1) and (2) argue validly:

A consideration can be a reason for action only if it is capable of motivating.

A consideration is capable of motivating only if it engages with the agent's present desires.

Therefore: A consideration can be a reason for action only if it engages with the agent's present desires.⁷

Followers of (3), though they accept the first premiss, do not go along with the second premiss, and therefore reject the above conclusion. Thus, the second premiss is obviously the root of contention between, on the one side, (1) and (2) and, on the other, (3).

Well, is it true that a consideration is capable of motivating only if it engages with the agent's present desires, as the disputed premiss claims? It might be claimed that a non-elliptical explanation of an intentional action always mentions the beliefs and the desires the agent had at the time of acting. This claim might be thought to indicate that the agent's present desires are a necessary component in the agent's motivation. Since this reasoning invokes our method of explaining action, let us call it, following Darwall, the Argument from Explanation.⁸

The best initial move to make in replying to this argument is to call attention to the distinction between explanatory (sometimes called motivating or effective) reasons and normative (sometimes called good, justifying, grounding, or recommending) reasons.⁹ To illustrate, suppose that you and I went to lunch with our boss, and that I ordered a certain dish whose taste I thought I would not like. This dish had a reputation for being a favourite among people with discriminating

taste, and I chose it because I thought this would impress the boss. Now let us add that the boss in fact believed that this dish was wildly overrated and that only a Philistine would choose it. Then, while the boss was away from the table you whispered to me, 'although I know the reason you ordered that dish, actually there was no good reason to order it'. Your statement looks like a straight contradiction. But it is not, because you were using 'reason' in two different senses: the first time in the explanatory sense and the second in the normative sense. The explanatory reason is the consideration that actually motivated me; but there was no normative reason for my action, because there was no consideration that in fact recommended it.

Now it might seem that those who believe that reasons for action are tied to present desires (or desires derived from present ones) are led to this belief by their failure to distinguish between normative and explanatory reasons.¹⁰ A reason that explains an intentional action must be one that engages with the motivations the agent had at the time of the action. Our theory of action bids us to accept this view about explanatory reasons, but it does not compel us to accept this view about normative reasons. That is, we can—without inconsistency—on the one hand explain action by reference to the beliefs and desires the agent had at the time of acting, and on the other hand believe that normative reasons are not tied to those desires.

We must be very careful, however, in what we infer from this. Invoking the distinction between explanatory and normative reasons can block the move from the theory of action to the

doctrine that all reasons for action have their basis in present desires. But, of course, blocking one way of supporting a doctrine is not the same as refuting the doctrine. In fact, we can make the distinction between explanatory and normative reasons while accepting the doctrine that all normative reasons for action come from present desires.¹¹ Indeed, this is demonstrated in the example I gave above to illustrate the distinction.

On the other hand, invoking the distinction between explanatory and normative reasons can help block a second way of supporting this doctrine. Some philosophers might say that ordinary English usage recognizes a distinction between 'A's reason for doing X' and 'the reason for A to do X'. They might then claim that the proper account of this linguistic distinction is that, roughly speaking, we use the expression 'A's reasons for doing X' to refer to reasons derivable from A's present desires, and the expression 'the reason for A to do X' to refer to the reason that we would act on if we were in A's situation (or, alternatively, reasons we hope A acts on¹²). They might say, 'This observation about usage reinforces the idea that we believe that reasons one has for doing something can only come from one's present desires. How else can we explain the difference in the two expressions?'

But this challenge can be easily answered. 'A's reason for doing X' refers to the consideration that, as it happens, motivates A to do X, whereas 'the reason for A to do X' refers to the normative reason for A to do X.¹³ Indeed, this explanation of the difference between the two expressions seems

at least as good as any other.

But I don't think there is really a significant difference in the meanings of the two expressions in the first place. Hence I shall continue to use 'A's reason for doing X' and 'the reason for A to do X' interchangeably. And, unless otherwise noted, I shall mean them both to be referring to normative reasons.

What other arguments are there for or against the disputed premiss that the only considerations capable of motivating are those that link up with the agent's present desires? One argument is not so much that that premiss is incorrect as that it is virtually empty. Suppose I describe an example in which someone comes to act benevolently. Catherine learns through the media about starving people in Africa and, though not previously benevolent, she joins a campaign to help these people. And suppose I say that, prior to learning about the suffering of these people, she did not have any desire to help people in distress. But in reply to this, adherents of the claim that the only considerations capable of motivating are those that plug into the agent's present desires, may (as I mentioned earlier in this chapter) agree that, if as result of the agent's confronting some new empirical fact or situation a new desire to do X springs up in the agent, he or she has a reason to do X. They would say that my example is therefore no counterexample to their theory. They would say that Catherine's confronting the empirical facts about the plight of the people in Africa causes a certain desire to arise in her, and this desire then gives her a reason to act. Critics of the claim that the

only considerations capable of motivating are ones which plug into the agent's present desires would reply that, rather, Catherine learns that she has a reason to help relieve the suffering, and discovering this reason then motivates her to act.

This last reply might seem to deny, however, the assumption on which the Argument from Explanation relies, the assumption that a non-elliptical explanation of intentional action would have to refer to the agent's beliefs and desires. If the reply essentially supposes that Catherine's new belief (that she has a reason to help relieve the suffering) on its own leads her to act, then it implies there is no need to mention a desire. And if there is no such need, we can throw out the assumption relied upon by the Argument from Explanation. Indeed, one way of objecting to the Argument from Explanation is to deny its account of explanation, that is, to deny that citing the agent's desires always does explanatory work. Nagel writes,

That I have the appropriate desire simply follows from the fact that these considerations motivate me; if the likelihood that an act will promote my future happiness motivates me to perform it now, then it is appropriate to ascribe to me a desire for my own future happiness. But nothing follows about the role of the desire as a condition contributing to the motivational efficacy of those considerations. It is a necessary condition of their efficacy to be sure, but only a logically necessary condition. It is not necessary either as a contributing influence, or as a causal condition.¹⁴

Then, the conclusion of this line of reasoning would be that, since the motivational work is being done by the consideration in question, the explanatory work must be done by citing this consideration, not the desire. Bond says Nagel thinks that 'beliefs can motivate on their own without aid of desires'

(though desires may be there as logical ghosts)'.¹⁵ If this is true, Nagel does indeed reject the belief plus desire model of rational explanation on which the Argument from Explanation relies. Bond, for his own part, strenuously denies that cognitions augmented by ghostly desires can move us to action. He maintains instead that beliefs about reasons create the relevant full-blooded desires, which then move the agent to action.

But, pace Bond, it is unclear whether Bond and Nagel really differ at a deep level. For, in the first place, Nagel himself mentions 'desires which an agent necessarily experiences in acting'.¹⁶ Furthermore, the contrast Nagel is chiefly interested in is the contrast between desires the agent is led to have by his or her other desires and desires the agent is led to have by his or her beliefs, and this is also, roughly, the crucial contrast for Bond.¹⁷ So I will assume that we can talk in terms of beliefs bringing about desires and then the two together causing action without offending against either Nagel's view or Bond's.

It should now be clear why we might think that the disputed premiss—viz., a consideration is capable of motivating only if it engages with the agent's present desires—cannot be used to argue for the claim that a consideration can be a (normative) reason only if it engages with the agent's present desires, where present desires are taken to be whatever desires the agent would have if he or she knew all the relevant empirical facts and was reasoning logically. Critics of this view will say to its defenders, 'You argue: since all reasons are

considerations capable of motivating, and all considerations capable of motivating engage with the agent's present desires, all reasons must be considerations that engage with the agent's present desires. Whether or not these three claims are acceptable depends on the term "present desires" is glossed. You admit that the motivations you have in mind are not merely the agent's desires before he or she learns the relevant true empirical facts, for you have granted that new desires can arise as a result of new empirical beliefs. And you acknowledge that the present desires you have in mind are those the agent would have after his or her reasoning has been corrected for logical mistakes. Perhaps you would also allow that the agent's antecedent desires can properly be adjusted in the light of some new evaluative information (for example, the agent, having an antecedent desire to act fairly, might, as a result of gaining an improved sense of what fairness is, acquire new particular desires). Well and good. But you don't allow the agent access to the most important information—namely, information about what there is reason to do. When you say that reasons are considerations capable of motivating, you mean capable of motivating after the agent's logic and empirical beliefs (and perhaps some evaluative beliefs) have been corrected. But when we say that reasons are considerations capable of motivating, we mean capable of motivating after the agent's beliefs about what there is reason to do have been corrected. And likewise, we would say that reasons are considerations that engage with the present desires the agent would have after learning what there is reason to do.'

Of course, those last remarks seem too circular to do much towards explaining reasons for action. Nevertheless, they are an effective rejoinder to the idea that because reasons for action must be capable of motivating, they derive from the desires the agent would have when he or she knows all the relevant empirical facts and is reasoning soundly—that is, they derive from the desires I am calling present desires.

SECTION 4

In this section I present an argument apparently showing that one can be morally required, and have a moral reason, to do an act even if one does not have a present desire that would be fulfilled by the act. Now (1) claimed that one is not morally required to do an act unless one has a present desire the act would fulfil. If the argument of this section is wholly successful, this claim is refuted. Both (1) and (2) claimed that one has no reason to do an act unless one has a present desire the act would fulfil. If my argument is wholly successful, this claim too is refuted.

Consider first the question of whether, from a self-interested point of view, an act is best only if one has some present desire the act would fulfil. Suppose that living to a ripe old age would enable Ms. Jones to have a better life, according to every plausible account of self-interest, than she would have if she lived a shorter life. Suppose further that she habitually smokes many cigarettes each day but gets very little pleasure from this habit. Let us also assume that smoking significantly reduces her chances of living to an old age, and

that giving up this habit would not cause her much discomfort. Then, it would be prudent of her to quit smoking. But what if she doesn't care about her own future welfare? Giving up smoking would nevertheless be best for her in the long run, for though present desires figure in self-interested calculation, they have no more weight (apart from their greater certainty) than future desires. In the light of these truths, could we plausibly deny that she has a prudential reason for giving up the habit? Doesn't it sound very odd to say, 'It would be prudent of her to quit, but she has no prudential reason for doing so'?¹⁸ And after accepting that she does have a prudential reason to quit, how could we deny that she has a reason to quit? Obviously, a prudential reason is a reason.

Having considered the prudential case, let us look at the moral case. Suppose that Mr. Smith tortures innocent children for fun. This activity is of course morally wrong whether or not Smith has any desire which would be fulfilled by his giving it up. Just as prudence's requirement that Ms. Jones do what will give her the best possible life taken as a whole is not undercut by her lack of concern for her own good, so morality's requirement that Mr. Smith not torture innocent children for fun is not undercut by his lack of aversion to this activity. Now, having granted that he is morally required to stop the torture, could we plausibly deny that he has a moral reason to stop? Is it not true that just as prudence generates prudential reasons, moral requirements generate moral reasons? And after accepting that he has a moral reason to stop, how could we deny that he does have a reason to stop? Obviously, a moral reason is a

reason.

Harman, champion of (1), would indeed say that Smith might have no reason, moral or otherwise, to stop torturing the children. On Harman's account, one has a moral reason to do X only if one accepts a moral convention requiring one to do X (where accepting a convention involves being motivated to act in conformity with it). According to this view, if one does not accept the convention and therefore does not have the corresponding desire, then one has no moral reason to do the act in question. And without a moral reason to do the act, one cannot (on this view) be morally required to do it. For Harman, if Smith does not accept any moral principle which forbids, or which implies a principle which forbids, torturing children for fun, then no such prohibition applies to Smith. That is, it cannot be appealed to in order to say that Smith is required or ought or has a reason to stop torturing the children.¹⁹

Harman believes this because he thinks that there are no moral principles that are 'more true' or 'better justified' than whatever principles are derived from the extant social conventions one accepts. For Harman, there are the moral conventions one person accepts, and there are the moral conventions another person accepts; but there is no single 'true' morality. He therefore thinks the appropriate principles to judge a person by are that person's own principles.²⁰

This view that moral requirements are determined by the agent's own present desires is extremely counter-intuitive. First of all, virtue and vice terms, e.g., 'dishonest', do apply to a person even if he or she does not care the least bit about

honesty or morality.²¹ But so do moral 'ought' judgements (and judgements about what the agent is morally required to do) apply whether or not the agent has the appropriate present desires.²² This is illustrated by the example about Smith torturing the children. We must therefore reject the claim, made by (1) and defended by Harman, that moral requirements are ultimately determined by the agent's present desires.

Of course, if the Smith example is completely convincing, not only is it the case that the agent can be morally required to do something without having any present desire the act would fulfil, but also there can be a normative reason for the agent to do something even if it will not fulfil any of his or her present desires. So if the Smith example is completely convincing, (2) should also be rejected, since it claims that normative reasons are relative to present desires. Since I do find the Smith example convincing, I reject (2). But there are some abstract considerations (as opposed to appeals to examples) which might seem to indicate that (2) is correct. I shall consider these in a later section.

SECTION 5

Having rejected both (1) and (2), I now ask which is the more plausible of the remaining views:

(3), the view that moral requirements provide not only normative reasons for action but also, for the objectively rational agent, motivations,

or

(4), the view that, although moral requirements do provide normative reasons for action, the agent can fail to be moved by these reasons without being in any way irrational.

Note that (3) and (4) agree that moral requirements generate reasons for action. The disagreement between these two views is over the connection between reasons for action and rational motivation. Which theory gets the better of this disagreement?

It may not yet be clear exactly what this disagreement is. So let us look more closely at the opposing views. Suppose someone asserts, 'I acknowledge that it is my moral duty to do such-and-such and that I therefore have a moral reason for doing it; nevertheless, I couldn't care less about doing it.' Adherents of (4) claim such an assertion is neither logically suspect nor indicative that the person is irrational, though he or she might well be psychologically abnormal. They would maintain that the desire to comply with one's reasons for action is merely a 'brute empirical fact'.²³ They might grant that this desire is so widespread that we have come to expect that someone who says 'I have a reason to do X' is motivated to some extent to do X. But they would insist that it is not irrational for the reasons one acknowledges not to be reflected in one's motivations. They would say that a lack of correlation between motivations and acknowledged reasons would spring from nothing more than a lack of some psychological propensity. It would be like, according to their view, having a tendency not to blink when bright lights flash. Such a tendency would be rare, but not for that reason open in any way to rational criticism.

Adherents of (3), on the other hand, would say that, by virtue of facts about our concepts, to believe that one has a reason to do X logically entails, for a fully rational person,

being motivated to some extent to do X. They would maintain that motivation (in the wide sense of 'motivation') which diverges from one's acknowledged reasons is rationally criticizable.

Those who accept (3) and those who accept (4) would agree that sometimes reasons and present motivations come apart. But (3)'s advocates would say, 'so much the worse for the rationality of those present motivations—since rational motivations are defined as those motivations that are in line with practical reasons'. Reacting quite differently, (4)'s advocates would say, 'of course motivations can diverge from one's reasons, but to say this is not to engage in rational criticism of those motivations—there is indeed no such thing as rational criticism of motivation'.

After we have agreed what reasons for action someone has, what does it matter whether we can call his or her motivation irrational? But this does seem, to me at least, to matter. Imagine that John believes that the strongest relevant normative reasons favour his doing X, and yet says that he is not motivated to do X. Suppose he claims that all we can say about him is merely that he is empirically unlike us, in that the things that he believes there are reasons for him to do are things he is just not motivated to do (in even the wide sense of 'motivation' in which any felt disposition to action counts as a motivation²⁴). But we do not merely believe that his motivation is empirically unlike ours—we pass the normative judgement that his motivation is rationally suspect. Indeed, there does seem to be a conceptual connection between reasons for action and rational motivation (in the wide sense of

'motivation').²⁵ What could rational motivation be if not motivation in line with the relevant practical reasons? So (3) rather than (4) seems to be the account we accept of the relation between normative reasons and motivation in the wide sense. Therefore, (3) is a more appealing view than (4).

SECTION 6

I have just now said that there is a conceptual connection between reasons for action and rational motivation. And my earlier example about Smith seemed to show that moral requirements yield reasons for action. But I do not think it would be plausible to claim simultaneously that moral requirements analytically yield reasons for action and that these reasons analytically yield motivation for the fully rational agent.

It is the claim that moral requirements generate reasons for action that I take not to be analytic. For, on the face of it, the proposition 'Morality as such does not give one reasons for action' is not self-contradictory.²⁶ And a closer look confirms this. Recall the broad and narrow senses of 'moral' and 'morality' ('ethical' and 'ethics') that I distinguished in Chapter 2. I said that in order for something to be one's morality in the broad sense it needs to fulfil only a single formal condition: that it be the set of action-guiding considerations one believes to be paramount. But if we interpret 'moral' in the broad sense, our question 'Why should I be moral?' becomes equivalent to 'Why should I abide by the values and ideals that I believe are paramount?' Since this question is

nonsensical, we must assume that someone who asks why he or she should be moral is using 'moral' in the narrow sense. I noted in Chapter 2 that the narrow sense of 'morality' is difficult to define. As I explained, some writers take morality—I return to shortening 'narrow morality' to 'morality'—to be distinguished by certain formal restrictions on its principles (other than the formal condition which distinguishes principles of broad morality); others say that morality can be defined (at least in part) by its social role; still others, that nothing useful can be said that is more general than that morality has to do with particular kinds of act, such as murder, assault, dishonesty, unfairness, ingratitude, and so on. Now according to all the plausible definitions of morality on offer, it is not true that, by virtue of the meaning of 'moral', moral requirements generate reasons, much less that they are overriding. Of course, someone might come up with a statement of the concept of morality which both is clearly superior to the alternatives and shows a straightforward conceptual connection with reasons for action. But this has not yet been accomplished.

I should mention the possible idea that it is not by virtue of the meaning of 'morality', but by virtue of the meaning of 'reason for action', that the connection between reasons and morality is analytic. However, I can't see how a plausible argument for this idea would go. So I shall ignore it.

As I indicated, perhaps the main reason for thinking that there is no simple conceptual connection between moral requirements and reasons for action is that someone who denies that moral requirements give reasons for action does not seem, on

the face of it, to be involved in self-contradiction. But perhaps there is some mistake here, though one more subtle than self-contradiction. One could be aware that morality requires one to keep one's promises and yet at the same time wonder whether one does have any reason to keep them, and one could do this without misunderstanding any of the words. Contrast the question, Does my being morally required to do something give me a reason for action? with, Does my being required by certain club rules to do something give me a reason for action? The second is certainly a sensible question, but is the first like the second? Indeed, club rules only sometimes provide reasons for action, this line of objection assures us. Such rules provide reasons only when there is some moral or prudential consideration favouring obedience of the club rule—when asked why a club rule yields a reason for action on a particular occasion, we cite some relevant moral or prudential consideration. But what happens when we ask why a moral requirement yields a reason? Appealing to morality itself seems ridiculously circular. Appealing to etiquette, club rules, chivalry, etc., seems backwards. Thus, it seems that morality is itself an ultimate source of reasons for action.

But it does not take much imagination to see what reply will immediately be thrown back at this line of reasoning. It will be that, though on first thought it seems we have reasons to do what conforms with (say) etiquette only when what conforms with etiquette happens to be what is required by morality or what serves self-interest, on further reflection we find that morality itself provides reasons only when it coincides with

self-interest—we find, in other words, that the ultimate practical considerations come from self-interest alone, not from self-interest and morality. Now this reply may ultimately be mistaken. But the reply comes too easily for the line of reasoning sketched in the previous paragraph to be convincing.

A different attempt to establish a conceptual connection between morality and reasons might be made by appealing to an analogy between practical and theoretical reason. There was a hint of the first part of this argument in the preceding chapter. As I noted there, we are interested in objective practical rationality. That is, we are interested in what acts and desires are rational, given all true beliefs; not merely in what acts and desires are rational, given the agent's beliefs. Thus we might be tempted by the following argument:

First premiss:	Objective practical rationality and objective theoretical rationality are analogous.
Second Premiss:	Objective theoretical rationality (as I explained in Chapter 7) involves only agent-neutral reasons for belief.
Intermediate conclusion:	Thus, objective practical rationality involves only agent-neutral reasons for action.
Additional premiss:	Consequentialism, a particular moral theory, is the only theory that confines itself to agent-neutral reasons for action.
Final conclusion:	Thus, this moral theory is the correct conception of objective practical rationality.

But, for a reason I mentioned in the previous chapter,

this argument is not compelling. It is vitiated by the weakness of its first premiss, there being a strong possibility of some relevant disanalogy between theoretical and practical rationality.²⁷ So the argument set-off above, built on a insecure premiss, fails to establish a conceptual tie between consequentialist morality and reasons for action.

Thus it seems we have no good argument for thinking that there is a conceptual connection. But we still have an argument for thinking there is not one—namely, that it is not self-contradictory to deny that moral requirements yield reasons. I therefore conclude that moral requirements do not analytically give reasons for action.

Some might think that if morality doesn't analytically give reasons, it doesn't give them at all. But this would, I think, be mistaken. Might it not be a substantive rather than analytic truth that morality generates reasons for action? The judgement that murder is wrong may well not be analytic and yet nevertheless be correct. Why can't the same be true of the judgement that moral requirements generate reasons for action? That is, the judgement that moral requirements generate reasons may have whatever status substantive judgements about right and wrong have.

To be sure, it may be unclear what this status is. There are well known problems with holding normative claims to be naturalistic, and yet if substantive normative propositions are neither conceptual nor naturalistic, what could they be? Or, is there not some way that we might hold, gropingly, that (pace G.E. Moore, other intuitionists, and non-cognitivists) the term

'reasons for action' stands for whatever natural considerations the optimal theory of reasons for action turns out to favour?

SECTION 7

Rather than speculating about answers to these questions, I turn now to some, mostly abstract, ideas that may appear to count against our accepting that moral requirements do provide reasons for action and that these reasons are sometimes stronger than conflicting reasons.

Consider first the idea that, however repugnant acting immorally may be, the mere fact that an act is immoral is not enough to make it irrational. As Philippa Foot has shown, the view that moral requirements generate reasons for action requires us to say that at least sometimes the mere fact that an act is immoral is enough to make it irrational.²⁸ Her argument is as follows. First she assumes, as I have been doing, that acting irrationally is doing something that is inconsistent with the reasons one has. Then she asks us to consider a case where the agent is morally required to do a particular act and has no reason of any kind for not doing it. If moral requirements do generate reasons for action, then not complying with the moral requirement is, in such a case, irrational. The upshot of this valid argument is that we cannot simultaneously hold both that immorality cannot make an act irrational and that morality provides reasons for action.

Foot's solution is to give up holding that morality provides reasons for action. But why not instead give up the idea that the mere fact that an act is immoral can^{not} be enough to

make it irrational? Well, one thing that may seem to favour Foot's solution is the idea that rationality consists merely in logical consistency. This idea is certainly correct with respect to subjective rationality. But, as I explained in the previous chapter, we are interested in objective rationality. We can too easily dismiss the claim that rationality consists merely in logical consistency as being true only of subjective rationality for it to constitute a convincing objection to the idea that moral requirements generate reasons.²⁹

Incidentally, suppose for the sake of argument that we ultimately conclude that (in the light of all true beliefs) moral requirements, properly construed, do yield the strongest reasons for action. Now as I have defined 'objective rationality', one's action is objectively rational only if it is in accordance with the strongest reasons for action. Our conclusion would thus be that immorality is objectively irrational, since acting immorally is not in line with the strongest reasons for action. But our coming to this conclusion would not commit us to the claim that there is something logically inconsistent about immoral actions. Assuming that objective rationality is more than mere logical consistency, we can accept the arguments of Parfit (aimed at, mostly, K. Baier) and Williams (aimed at A. Gewirth) that refusing to be guided by morality need not be logically inconsistent.³⁰

But let us turn now to the claim that there seems to be little point in calling the amoralist irrational. Foot thinks it seems silly to call a wicked, amoral person irrational, as if by this we 'can hit him where it will hurt'.³¹ And S. Blackburn

writes, 'we have quite enough invective at our disposal for people with horrid or queer desires.'³² He furthermore ridicules the attempt to find a 'knock-down argument that people who are nasty and unpleasant and motivated by the wrong things are above all unreasonable; They aren't just selfish or thoughtless or malignant or imprudent, but are reasoning badly, or out of touch with the facts. It must be an occupational hazard of professional thinkers to want to reduce all vices to this one.'³³ What can be said in reply to such jibes? First, we should grant that calling the amoralist irrational may not achieve much. But this is because the amoralist may give no weight to reasons for action, just as he or she gives no weight to moral considerations. 'Irrationalist' may not be a label which bothers such a person any more than 'amoralist'. But more important is this: the fact that calling amoralists irrational is pointless, if it is a fact, does not disprove that they are irrational.

Moreover, is there really no sense in saying that selfish and malignant people are 'reasoning badly' or 'out of touch with the facts'? Of course, Blackburn has in mind people who (knowing the relevant empirical facts and reasoning logically) do not care, or do not care enough, about morality. Granting this, however, we might still hold that they are out of touch with or insensitive to the fact that there is a moral reason to do the act in question.

Now it is tempting to think that the anti-realist about morality—e.g., Blackburn—will reply to this by saying that his anti-realism commits him to thinking that the only 'fact' in the

offing is that someone has a reason for doing a particular act only if he or she has a desire that would be fulfilled by the act. We might call this inference the idea that anti-realism about moral properties (e.g., wrongness) entails the Present-aim Theory about reasons for action. But this inference seems incorrect. The anti-realist can, according to Blackburn,³⁴ acknowledge that moral properties are not 'mind-dependent', that they would not necessarily alter if our attitudes altered, and that what makes it morally wrong (e.g.) to kick dogs is their pain, not our disapproval. But then an anti-realist could take the same line about reasons for action, saying, for example, that my reason not to prefer the destruction of the world to the scratching of my finger would not alter if I stopped caring about the world, or that what gives me a reason not to kick dogs is their pain, not my disapproval.

Indeed, anti-realists who want to accept the Present-aim Theory about reasons for action might seem to be caught in a dilemma. The first horn of the dilemma consists in accepting that, just as moral properties (such as the property of being morally required) are not mind-dependent, neither is the property of having a reason for action. Taking this horn involves giving up the Present-aim Theory. The other horn consists in accepting that, just as reasons for action are mind-dependent, so are moral properties (this is the view I called (1) at the beginning of this chapter). The problem for anti-realists with accepting this horn is that, as Blackburn recognizes, if anti-realism about moral properties entailed that they were mind-dependent, that would be a very strong objection

to anti-realism,³⁵ for it seems deeply mistaken to think that (e.g.) kicking dogs for fun would no longer be morally wrong if our attitudes and desires were different than they are now, or that Smith's torturing children for his own gratification is not wrong because he lacks any desire that would be fulfilled by his ceasing to torture them. So, assuming anti-realists face this dilemma, we would expect them to take the first horn, i.e. to accept that reasons are not mind-dependent. But if they do that, they should not jibe at the proposition that amoralists are mistaken about what reasons for action there are.

I admit, however, that some anti-realists might reject the above dilemma. They might do this by saying that it is an open question whether we must take the same line about reasons for action and that we take about moral properties. They are right—we cannot simply assume that the same line must be taken in both cases.

But in the anti-realists' escape from the dilemma I outlined, the prospect of a forceful anti-realist argument for the 'mind-dependence' of reasons for action (the Present-aim Theory) is lost. Such an argument would have to be built on the premiss that anti-realism about any (evaluative) properties, facts, or truths entails their mind-dependence. But anti-realists who reject the dilemma I outlined don't accept that premiss, since they claim that moral properties are not mind-dependent. This suggests a different dilemma: either anti-realism always entails mind-dependence, in which case this anti-realism is implausible since the property of (e.g.) moral wrongness is not mind-dependent; or anti-realism does not always

entail mind-dependence, in which case it does not commit us to the view that the property of having a reason for action is mind-dependent.

Having dealt with the anti-realist, I turn now to an objection, not to the claim that morality as such generates reasons, but to the further claim that moral reasons can be judged to be stronger than conflicting reasons. The objection is that because there is no neutral perspective from which to adjudicate between moral reasons and reasons of other kinds, we cannot know whether moral reasons always outweigh other reasons. But when people insist we must have a perspective to appeal to if we are going to be able to judge which kind of reason is stronger, what are they looking for? Perhaps a perspective defined by certain formal conditions which will then pick out one kind of reason for action as superior to other kinds. But, as far as I know, no attempt to argue for such a perspective has been convincing.

Nevertheless, we can weigh reasons of different kinds against one another. That is, we do judge that a reason of one kind happens to be stronger in this or that situation than a reason of another kind. Suppose you and I are lying in hospital beds in the same ward. No appeal to an abstract, super-normative perspective is needed to ascertain that, for example, your intense pain gives me a stronger (moral) reason to get an analgesic for you than the (self-interested) reason my slight thirst gives me to get a sip of water for myself.³⁶ (I shall discuss other examples in the next section.)

In this long section I have considered some fairly

abstract arguments suggesting either that morality doesn't generate (normative) reasons at all, or that even if morality does generate reasons, we can't judge its reasons to be stronger than other kinds. But, after scrutiny, none of these arguments seems cogent. So, buttressed by the earlier example about Smith, I maintain that moral requirements do generate normative reasons. And, buttressed by the recent example about us in hospital beds, I maintain that moral reasons can be (and can be judged to be) stronger than conflicting reasons.

SECTION 8

It is a further question whether moral reasons are always stronger than conflicting reasons. Consider an example of Foot's: I am faced with a choice between either humiliating someone in public or losing a large sum of my own money (which I would spend on frivolities anyway). About such cases she concludes, 'In the face of a sizeable financial consideration a small moral consideration often slips quietly out of sight.'³⁷ In reply, we might observe that, if this is an empirical observation about how people behave in such situations, what it shows about the normative question of which reasons for action are overriding is not clear. The normative question is 'Which reasons are strongest?', not 'Which reasons do people act upon?' However, Foot most likely means her statement as a frank reminder about what consideration we actually believe to be most important.

Well, has she presented us with a situation in which it is clear that a moral requirement provides a weaker reason

than that provided by self-interest? But we need to know more about the situation. For example, if morality does not require me to make a sacrifice for the benefit of someone who would not in fact make a similar sacrifice for my benefit if our roles were reversed, then it matters whether the other person would forgo a large sum of money in order to avoid humiliating me in public. Indeed, I may have good reasons for thinking it highly unlikely that this person would make the sacrifice if our roles were reversed.

This point about Foot's example suggests a more general point: often when people suspect that an agent has most reason to do what happens to conflict with some moral requirement, they are mistaken about what morality requires. Suppose I ask you a personal question I have no business asking. Some people might hold both that lying is always immoral and that you may nevertheless have more reason in this situation to lie than not to lie. These people would say that here we have an example of a moral reason's being outweighed by other reasons. But perhaps these people are mistaken about what morality requires of people who are asking such questions. Perhaps, as Sidgwick and Mackie hold, morality does not require honesty in these circumstances.³⁸ To cite a different example showing the same general point: many people believe both that certain sexual practices are wrong and that they have more reason to engage in these practices when they can get away with it than they have not to do so. Again, these people may be correct about what they have more reason to do, but incorrect about what morality requires.

The quickest way to induce scepticism in people about the supreme strength of moral reasons is to argue for a seemingly simple-minded moral code or for an extremely demanding one. As I have suggested, a deontological morality that is composed of simple rules, such as a rule that one never tell a lie, may well be an example. A different kind of example is standard Act-consequentialism, requiring one to do what one can to bring about the best consequences, each person's welfare being given equal weight. Suppose that I could give some large benefit to someone who is not connected to me in any special way, but that my doing this would involve a loss to me slightly smaller than the benefit given. As long as the benefit to the other person is larger than the loss to me, Act-consequentialism (according to most construals of that theory) says that it would be wrong of me not to make the sacrifice. But we may well suspect that the reason I have for not making the sacrifice is greater than the reason I have for making it.

In light of this kind of example, what should we think about the content of morality and the rational weight of its reasons? Some philosophers may answer that morality has the content the consequentialist conceives it as having, but that moral reasons as such might not (ever) have greater weight than some other reasons.³⁹ A different approach is to reject the view that morality requires one always to do what will generate the best consequences impartially considered. If this approach is taken, one option would be to embrace a deontological conception of morality. Then there would be the question of whether the conception should be a reciprocalist deontological

morality (which, as I have suggested, might require us not to injure, lie to, steal from, cheat, etc., only people who don't do these things to us) or a unilateralist deontological one (which would typically require us not to do these things even to people who do not show similar restraint in their dealings with us). Having a reciprocalist deontological morality would be, as I argued in Chapter 6, better from the point of view of self-interest than having a unilateralist one. Whether it is also the correct conception of morality is, however, a question too large to answer here.

But we may recoil from the demandingness of consequentialism in a quite different direction. We might, following S. Scheffler, temper consequentialism with an agent-relative 'prerogative' to devote one's resources (including time and energy) to one's own projects and welfare even when this will produce somewhat less good consequences viewed impartially.⁴⁰ Obviously, much more needs to be said in filling out this conception of morality, in particular much more about how much less good the consequences have to be before one is required to sacrifice one's own projects and welfare. And even if this were specified, there is, as S. Kagan has pointed out, another important problem with Scheffler's theory.⁴¹ Suppose the prerogative would permit the agent not to donate £10,000 to charity, although giving the money to charity would save lives in Africa. But, as Kagan complains, Scheffler's modified consequentialism would also appear to permit the agent to lie, cheat, steal, even murder, in order to get £10,000, and this is extremely counter-intuitive. Scheffler replies to this

objection by invoking the familiar consequentialist strategy of arguing that, since disastrous consequences would follow from our not having strong dispositions against such acts as lying, cheating, stealing, and causing injury, his theory would indeed tell us to have these dispositions.⁴² I cannot here anticipate the outcome of this debate. I only want to point out that, if Scheffler's theory is correct, there presumably are fewer counterexamples to the claim that moral reasons are overriding than there are if untempered consequentialism is correct.

But are there really no convincing counter-examples to the thesis that the reasons supplied by moral requirements are never weaker than conflicting reasons? Foot asserts that people often say they 'must' do an immoral act in order to avoid some personal disaster.⁴³ This sort of case seems to be different from the example discussed earlier where one is faced with either humiliating someone in public or losing a large sum of his money. (It is different because losing money which, by hypothesis, one would spend on frivolities should one not lose it would hardly count as a personal disaster.) So we need a new example. Suppose our country is being attacked by the army of a genocidal dictator. Five of us, connected to one another in no way except that we are fighting on the same side, discover that a large group of enemy soldiers is heading directly at us and is only a few fields away. We know that if we all stay and fight we shall all certainly be killed, and if we all flee we shall all probably be killed. But, luckily, there is another option: if one person stays to hold off the enemy, the other four of us will probably

reach safety. With this in mind, we all agree to draw straws to determine who stays behind to hold off the enemy, each of us swearing in good faith that he or she will comply with the result of the draw. Unfortunately for you, you draw the short straw.

After we have left, you nervously consider what to do. Fleeing gives you have some chance of escaping; staying gives you none. And (let us suppose) you know that if fleeing should get you safely behind our side's lines, no one there would ever find out about your broken agreement—because there is virtually no chance that any of the other four of us (who could reveal this) would also make it to safety. Given these facts, self-interest clearly bids you to turn and run as soon as the rest of us are far enough away not to threaten you.

Morality, however, demands that you stay and fight. To be sure, one may not be required, given normal circumstances, to uphold an agreement if this will lead to one's death. Most agreements might implicitly include an exception clause allowing any party to the agreement to break it if keeping it would be disastrous for him or her—i.e., most agreements might take the form: 'I agree to do such-and-such unless keeping the agreement will be disastrous for me'. But in the case we are considering, the agreement certainly did not include any such exception clause (in the circumstances, it would obviously have been a worthless agreement if it had). That is, the agreement each of us accepted was that the person drawing the short straw stays to hold off the enemy even though this will certainly result in his or her being killed. There seems, therefore, to be no good objection to the idea that morality requires you to keep

your side of this agreement in spite of the disastrous consequences for you. Any plausible conception of morality will, I believe, require you to stay and fight.⁴⁴

So you have a moral reason to stay and fight and a self-interested reason for fleeing to save yourself. Is this moral reason stronger than the self-interested one? It is difficult to insist that it is. Thus, the case we have been discussing may be a genuine counterexample to the thesis that the reasons yielded by moral requirements are always stronger than other reasons.

This is not to say, of course, that the reasons yielded by moral requirements never override self-preservation. Suppose you are in the same sad situation as before except that now the people with whom you have the agreement are your life-long friends, or even your spouse and children. These facts seem to strengthen your duty to keep your agreement.⁴⁵ And the moral reason for action would presumably be correspondingly stronger now—strong enough, it seems, even to override self-preservation.

SECTION 9

My conclusion, then, is that moral requirements as such do generate normative reasons for action, but (perhaps) not always reasons stronger than ones provided by self-interest. A further question is whether moral requirements ever generate weaker reasons than ones provided by present desires. I cannot think of an example in which they do. This is of a piece with my more general view that—contra Parfit—the rational challenge to morality comes from self-interest rather than from present

desires.

So the question 'Why be moral?' may admit of different answers in different circumstances. In most circumstances, the answer is that the relevant moral requirement gives you a stronger reason for action than you have for doing anything else. But in some circumstances the answer may be that, all things considered, you shouldn't be moral, since the self-interested reason you have in this case outweighs the moral reason.

NOTES TO CHAPTER 1

1. The question is, of course, at least as old as Plato's Republic.
2. 'Does Moral Philosophy Rest on a Mistake', in W. Sellars and J. Hospers, eds., Readings in Ethical Theory, Second edition, p. 86. This paper was originally published in Mind 21 (1912).
3. But cf. D. Parfit on the idea that being moral may be a mere means (Reasons and Persons, sect. 19).
4. Fifteen Sermons Preached at the Rolls Chapel, Sermon XI. And Hume echoed this sentiment when he asked, 'what theory of morals can ever serve any useful purpose, unless it can show, by a particular detail, that all the duties which it recommends, are also the true interest of each individual?' (An Enquiry Concerning the Principles of Morals, Section IX, Part II (p. 280 in P.H. Nidditch's revision of the Selby-Bigge edition))
5. For what I mean by benevolence, see the beginning of chapter 4 below.
6. As I explain in Chapter 8, I am talking about normative reasons, not about what are called explanatory or motivating reasons.
7. The titles 'the Self-interest Theory' and 'the Present-aim Theory' come from Parfit, op. cit., especially sects. 1, 45-6. Now Parfit describes a version of a 'Critical' version of the Present-aim Theory according to which (a) some desires are intrinsically irrational, (b) some desires are rationally required, (c) intransitivity of preferences is intrinsically irrational, and (d) given that I have the desires that are rationally required and lack all irrational desires and intransitivity, then I have most reason to do what will best fulfil my present desires (p. 119). And he mentions a possible version of the Critical Present-aim Theory which holds that having dominant moral desires is rationally required (pp. 119, 133). Parfit also mentions a different possible version of the Critical Present-aim Theory which holds a supreme concern for one's own good to be rationally required (p. 131). But I shall use the term 'Present-aim Theory' in such a way that these possibilities are excluded. In other words, Present-aim Theories, as I define them, never hold either that it is irrational not to care about self-interest or that it is irrational not to have overriding (or even any) moral desires. I make this qualification because I find that doing so makes formulating the issues easier.

NOTES TO CHAPTER 2

1. I mean by 'positive morality' roughly what R.B. Brandt means by 'social moral code'. Brandt thinks that, though there will be (in virtually any society) differences in individuals' moral beliefs, there will also be a large overlap. This overlap in moral beliefs is what Brandt defines as the 'social moral code' (A Theory of the Good and the Right, p. 172).
2. See P.F. Strawson, 'Social Morality and Individual Ideal', Freedom and Resentment, pp. 36, 40-41; G. Harman, 'Moral Relativism Defended', Philosophical Review, 84 (1975); The Nature of Morality, ch. 9; 'Relativistic Ethics: Morality as Politics', in P. French, T. Uehling Jr., H. Wettstein, eds., Midwest Studies in Philosophy, Vol. III: Studies in Ethical Theory, (1978), pp. 109-12; S. Lovibond, Realism and Imagination in Ethics, passim.
3. Methods of Ethics, Seventh edition, p. 30. Of course, Sidgwick's observation remains true (see Strawson, op. cit., p. 42; and D. Lyons, Ethics and the Rule of Law, p. 19). Sidgwick takes that observation to be a premise leading to the conclusion that 'no reflective person' identifies moral duty with what other people will disapprove of one for not doing (p. 30; see also p. 82-3; and Lyons, pp. 15-25).
4. It is not clear what this implies, for, as Harman says, 'Of course, a morality may rest on a tacit convention or agreement without the participants realizing it.' ('Relativistic Ethics', op. cit. in note 2 above, p. 117.)
5. Those who accept that morality must be rooted in a shared understanding are themselves eager to accept both that the individual might disagree with his or her society about what the society's shared moral ideas imply, and that the individual may well be correct. The reason these writers endorse this view is that it enables them to make room for criticism of (positive) morality. See Strawson, op. cit., pp. 39-40; Harman, 'Moral Relativism Defended', pp. 20-21, and The Nature of Morality, pp. 94, 95; Lovibond, op. cit., sections 24, 39-45.
6. J.L. Mackie, Ethics: Inventing Right and Wrong, p. 106. For examples of writers who distinguish between a wide and a narrow sense of 'moral' or 'ethical', though they would not all accept Mackie's account of the 'object' of narrow morality, see Sidgwick, pp. 25-6; W. Frankena, 'The Concept of Morality', in G. Wallace and A.D.M. Walker, eds., The Definition of Morality, pp. 156-8; Ethics, Second edition, pp. 6-7; and Thinking about Morality, pp. 19-28; J. McDowell, 'The Role of Eudaimonia in Aristotle's Ethics', in A. Rorty, ed., Essays on Aristotle's Ethics, pp. 365-6; S. Lukes, 'Taking Morality Seriously', in T.

Honderich, ed., Morality and Objectivity, pp. 98-9. And B. Williams writes,

It is possible to use the word 'ethical' of any scheme for living that would provide an intelligible answer to Socrates' question ['how should one live?']. In that sense, even the baldest egoism would be an ethical option. I do not think we should follow that use. However vague it may initially be, we have a conception of the ethical that understandably relates to us and our actions the demands, needs, claims, desires, and, generally, the lives of other people, and it is helpful to preserve this conception in what we are prepared to call an ethical consideration. (Ethics and the Limits of Philosophy, p. 12)

7. Cf. Williams, op. cit., p. 7 with p. 12.
8. See, for example, references in note 6 to Frankena's Ethics, to Mackie, to McDowell, and to Williams; and see P. Singer, Practical Ethics, p. 204.
9. H.L.A. Hart, The Concept of Law, pp. 169-176; Brandt, Theory of the Good and the Right, ch. 9 (especially p. 170); R.M. Hare, Moral Thinking, pp. 55-60; J. Rawls, A Theory of Justice, pp. 130-6, 149.
10. See (e.g.) K. Baier, The Moral Point of View, ch. 8; P. Foot, 'Moral Beliefs' and 'Moral Arguments', both reprinted in her Virtues and Vices; Frankena, 'The Concept of Morality'; G.J. Warnock, Contemporary Moral Philosophy, pp. 52-61. (But compare Williams, Morality: An Introduction to Ethics, pp. 87-8, 94-5.)
 Formalists might say that restrictions on content are a consequence of, rather than a competitor to, the formal conditions. (See Hart, op. cit., p. 176; Rawls, op. cit., p. 130 note 5; and Hare, op. cit., p. 63f.) Advocates of material conditions might respond by turning that claim on its head, asserting that the formal conditions are rather the consequence of the material conditions. Mackie, for example, suggests that moral principles must have certain formal features if they are to perform their function (Ethics, p. 152; see also his 'Morality and the Retributive Emotions', Persons and Values, Selected Papers vol. II, p. 217-8). But I shall not comment further on the relative priority of formal and material conditions.
11. See Hobbes, Leviathan, chs. XIII-XVII; Hume, A Treatise of Human Nature, Book III Part II. Cf. P. Nowell-Smith, Ethics, 1957, p. 201. See also Williams, Ethics and the Limits of Philosophy, pp. 7, 11, 12. On the social role of justice, see Rawls, op. cit., especially section 1, and p. 126; and 'Kantian Constructivism in Moral Theory', Journal of Philosophy, 77 (1980), p. 517.

12. See Mackie, Ethics, ch. 5; Hume's Moral Theory, chs. VI, VII, XI; and 'Co-operation, Competition, and Moral Philosophy', Persons and Values, especially pp. 157-164. (It is perhaps worth mentioning that theories which say that narrow morality is defined by its role or function in promoting interpersonal harmony would tend to exclude self-regarding behaviour from the purview of narrow morality.)
13. Sidgwick, p. 242.
14. As Neil Cooper seems to do in his articles 'Morality and Importance' and 'Two Concepts of Morality', both reprinted in Wallace and Walker, eds., The Definition of Morality.
15. See Sidgwick, p. 5; S. Toulmin, An Examination of the Place of Reason in Ethics, p. 202; Baier, op. cit., final section; Frankena, Ethics, p. 115; Thinking about Morality, p. 22; Singer, op. cit., p. 203; and S. Darwall, Impartial Reason, pp. 215-6.
16. Frankena defines teleological theories as those in which the morally right act is that which produces the greatest amount of non-moral goodness (Ethics, pp. 14-5). And T. Scanlon observes that 'standard consequentialist theories . . . generally resist the introduction of explicitly moral considerations into the maximand.' ('Rights, Goals, and Fairness' in S. Hampshire, ed., Public and Private Morality, p. 100)
17. Parfit, pp. 24, 26; S. Scheffler, The Rejection of Consequentialism, pp. 26-32; M. Slote, Common-sense Morality and Consequentialism, p. 36.
18. Williams, 'A Critique of Utilitarianism', in J.J.C. Smart and B. Williams, Utilitarianism: For and Against, pp. 86-9; A.K. Sen, 'Rights and Agency', Philosophy and Public Affairs, 11 (1982) pp. 28-33; Scheffler, op. cit., pp. 80, 84; Parfit, pp. 26-7.
19. For doubts about whether the anti-consequentialist would want to say this, see Foot, 'Morality, Action and Outcome', in Ted Honderich, ed., Morality and Objectivity, pp. 27-36; and 'Utilitarianism and the Virtues' Mind, XCIV (1985).
20. See references in note 18 to Sen and Scheffler; and also T. Nagel, Possibility of Altruism, ch. X (especially pp. 90-1); 'Subjective and Objective', Mortal Questions, p. 203-4; and 'The Limits of Objectivity', in S. McMurrin, ed., Tanner Lectures on Human Values, vol. I, pp. 102, 120; Mackie, Ethics, p. 157; Parfit, pp. 26-7, 143, 148. Much earlier than these authors, Hare discussed the idea that some moral principles contain 'agent-mentioning clauses' (Hare, 'Universalisability', reprinted in his Essays on the Moral Concepts, at p. 17, the original publication was in

1954-5 in the Aristotelian Society 55).

21. Mackie, Ethics, p. 157; Nagel, 'Limits of Objectivity', p. 102. (Note that agent-relativity does not preclude universality. Agent-relative principles contain individual variables, not individual constants. See Hare, 'Universalisability', pp. 16-18; and Moral Thinking, p. 140; and Nagel, Possibility of Altruism, pp. 48, 90f. Cf. Mackie, 'Sidgwick's Pessimism', Persons and Values, pp. 81-2, 87-8.
22. This is borrowed from Nagel's formulation of reason-stating propositions (Possibility of Altruism, pp. 47, 90-1).
23. Furthermore, if the moral 'ought' does imply 'can', perhaps all complete formulations of action-guiding moral principles must mention the agent in the antecedent as well. (See Parfit, p. 143.) That is, they take the form 'For all A and E, if A can promote E and if . . . , then A is required to promote E'. (The reference to the agent essential for agent-relativity is underlined.) Therefore, as I have defined agent-relativity, all complete action-guiding moral principles may perhaps be agent-relative, if 'ought' implies 'can'. But I shall ignore this kind of agent-relativity, and so henceforth omit the part of the antecedent clauses which would mention the agent's capacity to promote the event in question.
24. See Sidgwick, pp. 253, 261
25. See, again, Sidgwick, p. 242. C.D. Broad coined the useful term 'self-referential altruism' to refer to the special concern for the people connected to us in certain ways ('Moore's Ethical Doctrines' in P.A. Schilpp, ed., The Philosophy of G.E. Moore, pp. 53-5). Mackie says that 'Self-referential altruism forms, and always has formed, a large part of common sense morality.' ('Sidgwick's Pessimism', in op. cit. in note 21 above, p. 87).
26. Williams criticizes multi-level utilitarianism in a way that applies equally to multi-level consequentialism (Ethics and the Limits of Philosophy, pp. 106-110).
27. The ideas canvassed in the rest of this section come from Parfit, sections 10, 13. See also Sidgwick, pp. 434f.
28. Parfit, section 13.
29. Mackie writes, 'The happiness with which I am, inevitably, most concerned is my own, and next that of those who are in some way closely related to me.' (Ethics, p. 170. See also p. 132; and Mackie's 'Rights, Utility, and Universalization', Persons and Values, p. 203. Cf. Brandt, p. 218.) Apparently Mackie thinks that

impartial (equal) concern for all, even if this concern is weak, is psychologically impossible. Contrast Sidgwick's acknowledgement, 'it seems that most persons are only capable of strong affections towards a few human beings in certain close relations, especially the domestic' (p. 434, emphasis added). Sidgwick's remark, strictly interpreted, does not imply that most persons are incapable of impartial (equal) concern, but only that they are incapable of concern that is both impartial and strong (cf. Parfit, p. 27). Sidgwick's assumption, therefore, is weaker than Mackie's. I rely on that weaker assumption. Even I admit, however, that this assumption needs to be qualified. The assumption is about what is psychologically possible for a human being in the absence of some transformation through genetic engineering, neuroscientific techniques, or the like. (Some may think this qualification unnecessary, their argument being that no transformations could make us capable of having simultaneously deep attachments and equal concern, because deep attachments involve differential concern and thus are logically incompatible with equal concern. Slote writes, 'a person who really loves his child will in some circumstances favour that child rather than perform an act that would do more good, impersonally considered. . . . such a tendency to go against act-consequentialist optimality may actually be criterial, or definitional, of what it is to love someone.' (op. cit. in note 18, p. 93) But I do not need to take a stand on the correctness of Slote's claim.)

30. This distinction is made in these terms by G. Kavka ('The Reconciliation Project', in D. Copp and D. Zimmerman, eds., Morality, Reason and Truth: New Essays on the Foundations of Ethics, p. 302); but very much the same distinction can be found in Mackie, 'Co-operation, Competition, and Moral Philosophy', pp. 162-3; 'Norms and Dilemmas', Persons and Values, pp. 237-41; 'The Law of the Jungle: Moral Alternatives and Principles of Evolution', op. cit., pp. 130-1.
31. Kavka, loc. cit.

NOTES TO CHAPTER 3

1. Example of others who use the term 'prudence' in this way are: H. Sidgwick, Methods of Ethics, p. 381; R.M. Hare, 'Wrongness and Harm', Essays on the Moral Concepts, p. 103, and Moral Thinking, p. 190; P. Bricker, 'Prudence', Journal of Philosophy, 77 (1980), pp. 382, 401. (But others use 'prudence' in a different way, e.g., P. Foot, 'Are Moral Considerations Overriding?', Virtues and Vices, p. 187.)
2. Bishop Joseph Butler, Fifteen Sermons Preached at the Rolls Chapel, Sermon XI. (Cf. Hume, Enquiry Concerning the Principles of Morals, Appendix II.) See also B. Williams, Ethics and the Limits of Philosophy, p. 49.
3. Butler, Sermon XI. See also Sermon I, footnote 5.
4. Op. cit., sect. IX, part II (p. 282 in P.H. Nidditch's revision of the Selby-Bigge edition). See also sect. VI, part. I; sect. V, part. II; Conclusion, part II; and Appendix II.
5. For suggestions as to why this might have been only Hume's 'official' position, see D. Parfit, Reasons and Persons, p. 523-4.
6. To use Hume's famous line. (Enquiry Concerning the Principles of Morals, sect. IX, part I (p. 271 in the edition cited in note 4.)
7. A Theory of Justice, p. 568.
8. Ibid., p. 415, emphasis added. Sidgwick concurs: '. . . we can alter ourselves. . . . there may be within our power some process of training or hardening ourselves which may profoundly modify our sensibilities.' (p. 149) See also J. Glover, What Sort of People Should There Be? p. 88; and M. Midgley, 'The Objection to Systematic Humbug', Heart and Mind, pp. 82-4, 97-8.
9. D.A.J. Richards, A Theory of Reasons for Action, p.41; Rawls, op. cit., p. 415-6; Parfit, p. 8.
10. Glover, loc. cit. in note 8. (Here I have in mind self-transformations that no human being could effect. In later chapters I shall briefly discuss the possibility that some people may not be able to develop certain desires and dispositions that other people have developed or could develop.)
11. This definition of the theory is meant to follow Parfit's analysis (p. 487-490). See also J. Griffin's forthcoming book, chs. 1 and 2.

12. Parfit, pp. 496-499; Griffin, op. cit.
13. Blackburn writes, '. . . imagine someone whose only intention is that all his intentions get fulfilled? What does he do? But his problem arises because there is nothing else in the totality.' (Spreading the Word, p. 117)
14. Recall the view I discussed earlier that each person desires only that his or her life go as well as possible. The language of 'higher-order' and 'lower-order' can help explain why anyone would have held this view. As I explained, on most theories of self-interest, one's overall good is a higher-order desire, since this goal necessitates that one have other desires. Sex, food, affection, knowledge, and so on, are lower-order desires in the sense that they don't presuppose the having of other desires. According to Urmson, although Plato and Aristotle correctly perceived that one could not aim to promote one's overall good without simultaneously aiming to fulfil a lower-order desire, they made the mistake of thinking that the converse was also true. See J.O. Urmson, 'The Goals of Action', in A. Goldman and J. Kim, eds., Values and Morals, especially p. 139.
15. I shall later discuss the familiar objections to construing pleasure as a particular feeling or experience (or set of feelings or experiences).
16. This is a point Butler famously emphasized (op. cit., Sermon XI). (See also Hume, Enquiry Concerning the Principles of Morals, Appendix II, p. 301-2; Urmson, op. cit., pp. 138-141.) Sidgwick and Brandt accuse Butler of mistakenly holding that pleasures can come only from the gratification of antecedent desires (Sidgwick, pp. 44-5; Brandt, p. 84), and they urge us to leave open the possibility that one can get pleasure from the gratification of desires that did not exist prior to the time of their fulfilment.
17. Parfit, pp. 4, 466-7, 499.
18. What if some people objected that they each had only the single desire that their lives go as well as possible, and that according to their ascetic theory of personal good, being well off consisted in not having any desires at all? It would seem that their one desire could be achieved only by their not having any other desires. But we need to reject not these people's claims about what follows from their theory of personal good, but this theory of personal good itself. For convincing rejections of this sort of theory, see Aristotle, Nicomachean Ethics, 1152^bf.; and M. Slote, Goods and Virtues, pp. 131-9.
19. For suggestions along these lines, see W. Frankena, Ethics, second edition, ch. 5; and Parfit, pp. 501-2.

20. A few of the many possible examples are: Butler, Sermon XI; J. S. Mill, Autobiography, J. Stillinger, ed., pp. 85-6; Sidgwick, pp. 48, 403 (but cf. pp. 136, 138); B. Williams, 'A Critique of Utilitarianism', in Smart and Williams, Utilitarianism: For & Against, p. 112; Peter Singer, Practical Ethics, p. 218; J. Elster, Ulysses and The Sirens, p. 40; and Sour Grapes, p. 9; Parfit, sect. 2; P. Railton, 'Alienation, Consequentialism, and the Demands of Morality', Philosophy & Public Affairs, 13 (1984), p. 140.
21. Sermon XI.
22. Railton observes that 'one cannot be preoccupied with anything without this interfering with normal and appropriate patterns of thought and action.' (op. cit., p. 154)
23. For other examples meant to make the same point, see R.M. Adams, 'Motive Utilitarianism', Journal of Philosophy, 73 (1976), p. 471; and Parfit, sect. 2.
24. I am following Adams and Parfit in holding that the value of a motive or set of motives may be greater than the value of the actions those motives lead one to perform. See Adams, op. cit., p. 470; Parfit, p. 472.
25. Ethics: Inventing Right and Wrong, p. 192. See also p. 188.
26. Cf. Sidgwick (pp. 47-50); Harman, Nature of Morality, p. 149; and Railton (op. cit. in note 20, p. 144). In addition, see Harman's discussion of the phenomenon of 'stimulus generalization' of the concern and respect that one develops for certain others ('Justice and Moral Bargaining', Social Philosophy and Policy, 1 (1983), p. 125-6).

NOTES TO CHAPTER 4

1. R.B. Brandt, 'The Psychology of Benevolence and Its Implications for Philosophy', Journal of Philosophy, 73 (1976), pp. 430-433; and A Theory of the Good and the Right, p. 138.
2. This is one of the reasons that some consequentialists might hold that such intense sympathy may not always be morally optimal.
3. See A.K. Sen, 'Rational Fools', in F. Hahn and M. Hollis, eds., Philosophy and Economic Theory; and especially M. Overvold, 'Self-Interest and the Concept of Self-Sacrifice', Canadian Journal of Philosophy, 10 (1980); and 'Self-Interest and Getting What You Want', in H. Miller and W. Williams, eds., The Limits of Utilitarianism.
4. H. Sidgwick, Methods of Ethics, pp. 502-3; cf. P. Singer, Practical Ethics, p. 213. My claim is meant to be weaker than Aristotle's opinion that 'without friends no one would choose to live, though he had all other goods'. (Nicomachean Ethics, 1155^a) See also 1169^b, and M. Slote, Goods and Virtues, pp. 133-5.
5. Singer hints that psychopaths may not necessarily be mentally sick or even unhappy. (Practical Ethics, pp. 214-16) Nevertheless, they are typically very imprudent, in that they typically either give less weight to their future interests than they do to their present interests, or fail even to consider the long-term consequences of their present actions. Psychopaths thus do not show much concern either for other people or for their own further future. The latter characteristic often ruins their lives. But does the former? Maybe their indifference to their further future often ruins their lives before their indifference to others can.
6. Perhaps it is easy to give an evolutionary explanation of why this is so. But that is not my concern here.
7. The pretence of having concern for others is not the only pretence difficult to keep up for a long period. In the next chapter I shall discuss the difficulty of sustaining other pretences for long periods.
8. G. Ryle, 'Pleasure', in J. Feinberg, ed., Moral Concepts; Brandt, Theory of the Good and the Right, pp. 36-8; D. Parfit, Reasons and Persons, p. 493.
9. This view is explained in Brandt, loc. cit. And the criticisms of this view that I rehearse in the text are Brandt's.
10. Hare, 'Pain and Evil', in Feinberg ed., Moral Concepts;

and Moral Thinking, p. 93; Brandt, Theory of the Good and the Right, pp. 37, 131-2; Parfit, pp. 493, 501.

11. For other endorsements of the view that what is common to all pleasures is only that they are experiences that are connected with our desires in a particular way, see Sidgwick, p. 127 and the references cited in the preceding note.
12. Sidgwick, pp. 434, 502; J.L. Mackie, Ethics, p. 170.
13. Especially Hobbes, Leviathan, chs. XIII-XVII; Hume, A Treatise of Human Nature, bk. III part. II. For a recent interesting complement to these arguments, see R. Axelrod, The Evolution of Cooperation, especially p. 87.

NOTES TO CHAPTER 5

1. The Moral Point of View, p. 309.
2. This point can be put abstractly as follows: although unilateral adherence to the moral rules is better for each person than mutual violation, unilateral violation is even better for each person than mutual adherence. For this way of putting the point, I am indebted to D. Gauthier, 'Thomas Hobbes: Moral Theorist', Journal of Philosophy 76 (1979), p. 555.
3. Of course this doesn't mean they will treat well those people caught being immoral, especially if this immorality was towards them.
4. For a good discussion of the Hobbesian argument, see S. Darwall, Impartial Reason, ch. 13.
5. That there have been so many unsolved serious crimes indicates that people have often gotten away with the mere appearance of conformity with the most important norms of the positive morality.
6. The importance of this distinction in the present context is argued by G. Kavka, 'The Reconciliation Project', D. Copp and D. Zimmerman, eds., Morality, Reason and Truth, p. 301.
7. J. Hospers, Human Conduct: An Introduction to the Problems of Ethics, pp. 180-1, quoted in K. Nielsen, 'Why Should I Be Moral?', in W. Sellars and J. Hospers, eds., Readings in Ethical Theory, Second edition, p. 767.
8. Perhaps J. Rawls's name first comes to mind here, but Rawls admits 'the maximin rule is not, in general, a suitable guide for choices under uncertainty.' (A Theory of Justice, p. 153) He claims merely that maximin is appropriate when in a situation having certain features (beyond uncertainty about probabilities). But for an argument that something like maximin is the appropriate strategy for prudential choice under uncertainty, see Kavka, op. cit., p. 301.
9. K. Arrow and L. Hurwicz, 'An Optimality Criterion for Decision Making Under Ignorance', in C. Carter and J. Ford, eds., Uncertainty and Expectations in Economics: Essays in Honour of G.L.S. Shackle, pp. 1-11.
10. Some might worry that there is a danger of infinite regress here. The agent might start out with information about the probabilities of the alternative outcomes of some prospective act; and then, stepping back, have thoughts about how reliable that information is; and then, stepping back a

second time, have thoughts about his or her ability to assess the reliability of the information; and then, stepping back a third time, ... But I think we can put aside this worry about infinite regress: it seems that the appropriate number of times the agent needs to step back is just two.

11. Concerning this bias, see D. Parfit, Reasons and Persons, p. 124.
12. Hume, Enquiry Concerning the Principles of Morals, sect. IX part II, (in Selby-Bigge's edition, p. 283).
13. See Hare, Moral Thinking, p. 195. Cf. D.A.J. Richards, A Theory of Reasons for Action, p. 290.
14. This point is noted by G. Harman (The Nature of Morality, p. 148).
15. See P. Foot, 'Moral Beliefs', Virtues and Vices, p. 129. See also P. Singer, Practical Ethics, p. 213.
16. Kavka, op. cit., p. 306. In support of the claim that acting morally can bring the moral person pleasures, see (e.g.) H. Sidgwick, Methods of Ethics, p. 150; Parfit, p. 30; Mark Overvold, 'Self-interest and Getting What You Want', in H. Miller and W. Williams, eds., The Limits of Utilitarianism, p. 191.
17. And some people would think less of others for not doing so. See Aristotle, Nicomachean Ethics, 1121^a2.
18. Cf. the beginning of the previous chapter for the discussion of how acting on benevolent desires can make one worse off.
19. Kavka, p. 318 note 16.
20. Sidgwick, pp. 174-5.

NOTES TO CHAPTER 6

1. See Mackie's qualified acceptance of universalizability, Ethics, ch. 4, and 'The Three Stages of Universalization', Persons and Values, Selected Papers, vol. II. Many philosophers accept that what is morally right for one person is morally right for any other relevantly similar person in similar circumstances, and yet believe that universalizability, even in conjunction with prescriptivity, will not necessarily lead everyone to the same substantive principles: e.g., Sidgwick, pp. xix, 209-210, 379; W. Frankena, Ethics, Second edition, pp. 25, 38; A. Quinton, Utilitarian Ethics, p. 68; T. Nagel, 'The Excessive Demands of Impartiality', London Review of Books, 1-14 July 1982; B. Williams, Ethics and the Limits of Philosophy, pp. 36-8, 66, 82, 115-6. Rawls also accepts that moral principles must be universalizable, though he calls this feature of principles 'generality' (Theory of Justice, pp. 131, 132). For a famous attack on universalizability, see P. Winch, 'The universalizability of moral judgments', The Monist 49 (1965).
2. The formulations of the two kinds of Egoism come from Rawls (op. cit., p. 124). Rawls (pp. 135-6) suggests that these two forms of Egoism are morally unacceptable for the reason I mention in the text. Mackie does not distinguish between these two kinds of Egoism, but it is clear that he would consider both of them ruled out by the least problematic 'first stage of universalization' (Ethics p. 84; see also Mackie's 'Sidgwick's Pessimism', Persons and Values, pp. 81-2, and his 'The Three Stages of Universalization', p. 172 of the same volume).
3. Methods of Ethics, p. 87. See also R.B. Brandt, A Theory of the Good and the Right, p. 276. J. Glover, who is more sympathetic than Brandt to act-utilitarianism, acknowledges that 'Living up to this demanding morality is to many of us a very unattractive prospect: it is the prospect of a huge reduction in income and the loss of a lot of our spare time.' (Causing Death and Saving Lives, p. 110) See also Williams, Ethics and the Limits of Philosophy, p. 77.
4. 'Justice and Moral Bargaining', Social Philosophy & Policy, 1 (1983), p. 124. (Cf. Harman's The Nature of Morality, p. 111, and his 'Relativistic Ethics: Morality as Politics', Midwest Studies in Philosophy 3 (1980), p. 114.) S. Scheffler identifies this as one of the causes of our persistent loyalty to deontological conceptions. He writes, 'The rejection of agent-centred restrictions accomplishes a redistribution of moral protection, and one source of our reluctance to reject the restrictions may be a sense that we are among those with the most to lose in such a

redistribution.' (The Rejection of Consequentialism, p. 113)

5. Mackie, Ethics, p. 191.
6. Sidgwick claimed that it is (p. 215); cf. Parfit, p. 98/9.
7. On Common-sense Morality's imprecision, see Sidgwick, bk. III, especially ch. XI.
8. Indeed Robert Axelrod shows that, in a round-robin in which each player plays a series of two-person prisoner's dilemma games against each other player, a strategy based on reciprocity (he calls it 'tit-for-tat') does better in self-interested terms than a very wide variety of other strategies. In such a tournament, tit-for-tat comes out ahead of both, on the one hand, more forgiving and altruistic strategies and, on the other, more punitive and selfish strategies (The Evolution of Cooperation, Part. II). A further interesting point, made by both Axelrod and Mackie, is that in a world pervaded by partial-conflict situations, the presence of 'unilateral altruists' (people who are kind to others who neither reciprocate nor make their fair contribution to public goods) makes precarious their own survival as well as the survival of 'reciprocal altruists' (Axelrod, pp. 136-9; Mackie, 'The Law of the Jungle: Moral Alternatives and Principles of Evolution', Persons and Values, pp. 127, 131).
9. See, in Mackie's Persons and Values, 'Co-operation, Competition, and Moral Philosophy', pp. 163-5; 'Norms and Dilemmas', pp. 239-41; and especially 'Law of the Jungle', pp. 30-1. Kavka takes a similar view (op. cit., p. 302), and cf. Hobbes's first two laws of nature (Leviathan, ch. XIV).
10. Kavka, op. cit. pp. 312-3. See also references to Harman cited in note 4 above.
11. See Chapter 2, Section 2 and corresponding notes.

NOTES TO CHAPTER 7

1. D. Parfit, Reasons and Persons, Part Two.
2. *Ibid.*, Part One sects. 1-9, 32, and Part Two.
3. Parfit discusses and rejects this theory (p. 23). D. Gauthier, on the other hand, explicitly endorses it. He writes, 'It may be tempting to suppose that it is rational to form an intention if and only if it would be utility maximizing to execute the intention. Instead we argue that it is rational to execute an intention if and only if it is utility maximizing to form it.' ('Deterrence, Maximization, and Rationality', Ethics, 94 (1984), p. 483)
4. Parfit, sect. 34. (B. Williams writes, 'It is a grotesque product of theory and strenuous moralism to suppose that "moral" and "prudential" sufficiently divide up the justifiable motives or reasons for doing something....' (Morality: An Introduction to Ethics, p. 85))
5. *Op. cit.*, sects. 35, 45.
7. ~~Parfit~~, sects. 45-6.
6. For an elaborate presentation of the Deliberative Present-aim Theory, where these ends are criticized, see R. Brandt, A Theory of the Good and the Right, especially pp. 116-126.
8. This example is inspired by Anscombe (Intention, sect. 37). But Anscombe's claim is that one simply cannot want a saucer of mud, just as such.
9. Rawls, A Theory of Justice, p. 432. Rawls does not himself accept that the preference for counting blades of grass would be irrational, though he does seem to think that it is an empirical fact that few if any of us would have such a preference. But I am taking the preference to count blades of grass to be a possible example of an irrational preference.
10. Parfit gives the examples of 'Tuesday-indifference' and 'Within-a-Mile Altruism' (sect. 47).
11. Parfit, sects. 52, 72.
12. *Ibid.*, pp. 27, 143.
13. Nagel, The Possibility of Altruism, chs. VI and VII; Parfit, sect. 35.
14. Parfit, p. 148.

15. Ibid., ch. 7.
16. Ibid., p. 144. (Note that Parfit does not claim that this formal argument totally refutes the Direct Self-interest Theory; he claims merely that this theory suffers serious damage at the hands of that argument.)
17. Ibid., p. 191.
18. Ibid.
19. See Williams, 'Ethical Consistency', in Problems of the Self; and 'Conflicts of Values', in Moral Luck.
20. Williams insists on a disanalogy between 'reflective factual deliberation' and 'reflective practical deliberation' Ethics and the Limits of Philosophy, pp. 66-9. I come back to this later (ch. 8 sect. 6).
21. Parfit, pp. 25, 153.
22. Ibid., p. 191.
23. Cf. Parfit, pp. 126, 130, 192.
24. Ibid., sect. 50; see also sect. 60.
25. We might be tempted to think that the Direct Self-interest Theory discriminates between rational and irrational desires on the basis of their content. The temptation is to think that the Direct Self-interest Theory would say that desires for things for oneself (such as desires for a drink, for a sexual encounter, for power of some kind, or for a particular personal achievement) are rational, but desires for things for other people (such as altruistic desires) are irrational. But this is a mistake. For, as Butler reminds us (see references in my notes to ch. 3), desires like thirst for alcohol, lust, ambition, and so on, do often lead people to do what will have, for them, worse consequences in the long run. Though the object of each of these desires is something for the agent, acting on these desires in such cases is, according to the Direct Self-interest Theory, irrational. And the Direct Self-interest Theory has to say that caring more about anything, even something for oneself, than about long-term self-interest is irrational. So what this theory claims are irrational are not particular desires, but patterns of desire in which self-interest sometimes plays a subordinate role.
26. Schelling, The Strategy of Conflict, esp. p. 36; Parfit, sects. 5 and 8, and Appendix A.
27. Strategy of Conflict, p. 36.

28. Parfit, sect. 9.
29. Hare, Moral Thinking, pp. 48-52, 130-140. It should be noted that even if utilitarianism is ultimately unacceptable, this would not mean that Hare's kind of argument has no force.
30. Parfit, p. 130.
31. *Ibid.*, p. 467.

NOTES TO CHAPTER 8

1. G. Harman, The Nature of Morality, chs. 8-11; 'Relativistic Ethics: Morality as Politics', Midwest Studies in Philosophy III (1978), pp. 111-112; 'Justice and Moral Bargaining', Social Philosophy and Policy, 1 (1983), especially p. 121.
2. B. Williams sometimes seems to be an exponent of (2). See 'Internal and External Reasons', in Moral Luck, especially pp. 110, 112-3. (Cf. P. Singer, Practical Ethics, p. 207.) And for a view that differs from (2) only in thinking that 'interests' as well as desires yield reasons for action, see Philippa Foot, 'Reasons for action and desires', in Virtues and Vices, pp. 152-3, 156; 'Morality as a System of Hypothetical Imperatives', in op. cit., especially pp. 172-3 (but cf. 'A Reply to Professor Frankena', in op. cit., p. 179).
3. Harman, Nature of Morality, pp. 125-131; B. Williams, 'Internal and External Reasons', pp. 104-5; R.B. Brandt, A Theory of the Good and the Right, ch. VI.
4. See, for example, H. Sidgwick, Methods of Ethics, seventh edition, p. 34; J. McDowell, 'Non-Cognitivism and Rule-Following', in S. Holtzman and C. Leich, eds., Wittgenstein: To Follow a Rule, pp. 154-5, 161 notes 19, 20; T. Nagel, The Possibility of Altruism, pp. 8-9; 'The Limits of Objectivity', in S. McMurrin, ed., Tanner Lectures on Human Values, vol. I, p. 110.
5. D.A.J. Richards, A Theory of Reasons for Action, pp. 64, 89, 242-3.
6. With regard to (1), see Harman, Nature of Morality, p. 128; and his 'Is There a Single True Morality', in D. Copp and D. Zimmerman, eds., Morality, Reason and Truth, p. 36. With regard to (2), see Williams, 'Internal and External Reasons', op. cit., p. 106. With regard to (3), see, e.g., Nagel, Possibility of Altruism, p. 32.
7. Cf. the picture given in Harman, 'Is There a Single True Morality?', p. 38. Barry Stroud says that the principle 'reasoning produces a want or propensity only if the agent also has some prior want or propensity' is a plausible one (Hume, p. 170).
8. S. Darwall, Impartial Reason, p. 36.
9. W. Frankena, 'Obligation and Motivation in Recent Moral Philosophy', in W. Sellers and J. Hospers, eds., Readings in Ethical Theory, second edition, p. 710; J. Raz, 'Introduction', in Raz, ed., Practical Reasoning, p. 2;

- E.J. Bond, Reason and Value, ch. 2; Darwall, op. cit., pp. 28-34.
10. See the references to Frankena and Darwall in the previous note.
 11. Bond shows why we must make a distinction between motivating and grounding reasons even if we accept the view that all of one's reasons for action are tied to one's antecedent desires (p. 28). But making this distinction does not force us to give up this view. On the other hand, Bond thinks, 'The confusion between the two kinds of reasons leads to the supposition that somehow grounding reasons must be tied to wants' (p. 40, emphasis added) In fact, quite a lot of Bond's argument against the doctrine that grounding reasons are always based in antecedent desires consists of invoking this distinction.
 12. Harman prefers this (to my mind inferior) alternative. See his 'Relativistic Ethics', Midwest Studies III, p. 112.
 13. See Darwall, Impartial Reason, p. 32. But compare Williams, Ethics and the Limits of Philosophy, p. 192.
 14. Nagel, Possibility of Altruism, pp. 29-30. See also McDowell, 'Are Moral Requirements Hypothetical Imperatives?', Proceedings of the Aristotelian Society, suppl. vol. 52 (1978), pp. 15-16; and 'Non-Cognitivism and Rule-Following', in Holtzman and Leich, eds., pp. 154-5.
 15. Bond, op. cit., p. 71. See also pp. 11-14.
 16. Possibility of Altruism, pp. 30-2.
 17. Nagel, *ibid.*; Bond, op. cit., pp. 13, 71.
 18. Indeed, G.R. Grice insists that the proposition, 'It is in my interest to do X, but that is no reason for my doing X' is self-contradictory (The Grounds of Moral Judgement, pp. 18, reprinted in Raz, ed., Practical Reason, p. 175). By contrast, J. McDowell claims that although prudential considerations do provide reasons for action, they do not provide reasons for action if the prudential considerations conflict with what is required by moral virtue ('Are Moral Requirements Hypothetical Imperatives?', pp. 25-9).
 19. Harman writes, 'In my view, the basic moral principles that apply to a given person are principles that person actually accepts as principles regulating the conduct of members of some group to which the person belongs' ('Justice and Moral Bargaining', Social Philosophy and Policy, 1 (1983) p. 123) See also Nature of Morality, pp. 109-10; 'Relativistic Ethics', Midwest Studies, III, p. 110; and

- 'Is There a Single True Morality', in Copp and Zimmerman, eds., p. 36.
20. This is clearly implied on the top of page 84 of Nature of Morality. And see 'Is There a Single True Morality', pp. 38-9.
21. See Foot, 'Morality as a System of Hypothetical Imperatives', Virtues and Vices, p. 172. Williams concurs: 'There are of course many things that a speaker may say to one who is not disposed to G when the speaker thinks that he should be, as that he is inconsiderate, or cruel, or selfish, or imprudent: or that things and he, would be a lot nicer if he were so motivated.' ('Internal and external reasons', Moral Luck, p. 110) Harman would accept this, but he calls such judgements 'evaluative' rather than 'moral' (Nature of Morality, p. 108).
22. Foot, *ibid.*, p. 159.
23. Richards, Theory of Reasons for Action, pp. 64f., 242f.
24. This formulation of the wide sense of motivation comes from R.M. Hare, Freedom and Reason, p. 170.
25. Nagel writes, 'I contend that the judgment that one has reason to do something includes the acceptance of a justification for doing it, and that this is its motivational content.' (Possibility of Altruism, p. 65.) Bond claims that it is 'senseless to say of someone that he acknowledges something as a reason if he is not, to some extent, motivated towards doing the thing...' (*op. cit.*, p. 33; and see pp. 5, 40, 41, 69, 159) See also Darwall, *op. cit.*, pp. 19, 31, 51; and cf. D. Davidson, 'How is Weakness of the Will Possible?', in Essays on Actions and Events, p. 27.
26. This is also true, pace Grice (see note 18 above), about the proposition that self-interest does not generate normative reasons for action.
27. Williams claims that whereas reflective factual deliberation is 'not essentially first-personal', practical deliberation—even reflective practical deliberation—is so (Ethics and the Limits of Philosophy, pp. 66-9). We should admit, I think, that this claim rings true, though Williams's latest attempt (*ibid.* p. 69) to provide support for it is of little real help. See effective criticism by P. Seabright, 'Character', London Review of Books, 5 September 1985, p. 18.
28. Foot, 'Reasons for Action and Desires', Virtues and Vices, p. 153.

29. For examples of writers who do not identify practical rationality with mere logical consistency, see R. Norman, Reasons for Action, p. 64; Nagel, 'The Limits of Objectivity', op. cit., pp. 104-5; Parfit, Reasons and Persons, sect. 46.
30. Parfit, sect. 33; Williams, Ethics and the Limits of Philosophy, ch. 4.
31. Foot, 'Reasons for Action and Desires', p. 152.
32. 'Rule-following and moral realism', in Holtzman and Leich, eds., Wittgenstein: To Follow a Rule, p. 181.
33. Spreading the Word, p. 222.
34. 'Rule-following and moral realism', p. 179; Spreading the Word, pp. 217-219; 'Errors and the Phenomenology of Value', in T. Honderich, ed., Morality and Objectivity, p. 6. ('Mind-dependence' may have more than one meaning, but I shall use the term with the meaning Blackburn gives it.)
35. See references in the previous note.
36. Nagel, 'The Limits of Objectivity', pp. 109-110.
37. 'Are Moral Considerations Overriding?', Virtues and Vices, p. 184.
38. Sidgwick, p. 318; J.L. Mackie, Ethics: Inventing Right and Wrong, p. 183.
39. Sidgwick is perhaps the most obvious example ('Concluding Chapter', especially pp. 498, 508). See also P. Singer, Practical Ethics, pp. 204-8, 219-20.
40. The Rejection of Consequentialism, passim, but especially pp. 22 and 96-8.
41. 'Does Consequentialism Demand Too Much? Recent Work on the Limits of Obligation', Philosophy and Public Affairs, 13 (1984), p. 251.
42. Scheffler presented this reply in a lecture in Oxford in June, 1985.
43. Foot, 'Are Moral Considerations Overriding', p. 185.
44. That even Scheffler's theory would give this verdict can be surmised from his acknowledgement: 'when one makes a promise ... one ordinarily forfeits the prerogative to give one's own interests more weight than the relevant interests of those to whom one becomes bound.' (op. cit., p. 23 note 8)

45. J. Rawls writes, 'deceit and infidelity are always wrong But they are not always equally wrong. They are worse whenever bonds of affection and good faith have been formed' (A Theory of Justice, p. 475/6)

LIST OF WORKS CITED

- Adams, R.M., 'Motive Utilitarianism', Journal of Philosophy, 73 (1976).
- Anscombe, G.E.M., Intention, Second edition, Oxford: Blackwell, 1963.
- Aristotle, Nicomachean Ethics.
- Arrow, K., and Hurwicz, L., 'An Optimality Criterion for Decision Making Under Ignorance', in C. Carter and J. Ford, eds., Uncertainty and Expectations in Economics: Essays in Honour of G.L.S. Shackle, Oxford: Blackwell, 1972.
- Axelrod, R., The Evolution of Cooperation, New York: Basic Books, 1984.
- Baier, K., The Moral Point of View, Ithaca, New York: Cornell University Press, 1958.
- Blackburn, S., Spreading the Word, Oxford: Oxford University Press, 1984.
- 'Rule-following and moral realism', in S. Holtzman and C. Leich, eds., Wittgenstein: To Follow a Rule, London: Routledge & Kegan Paul, 1981.
- Bond, E.J., Reason and Value, Cambridge: Cambridge University Press, 1983.
- 'Errors and the Phenomenology of Value', in T. Honderich, ed., Morality and Objectivity, London: Routledge & Kegan Paul, 1985.
- Brandt, R.B., A Theory of the Good and the Right, Oxford: Oxford University Press, 1979.
- 'The Psychology of Benevolence and Its Implications for Philosophy', Journal of Philosophy, 73 (1976),
- Bricker, P., 'Prudence', Journal of Philosophy, 77 (1980).
- Broad, C.D., 'Moore's Ethical Doctrines' in P.A. Schilpp, ed., The Philosophy of G.E. Moore, La Salle: Open Court Publishing Co., 1942.
- Butler, Bishop J., Fifteen Sermons Preached at the Rolls Chapel, 1726.
- Cooper, N., 'Morality and Importance', in G. Wallace and A.D.M. Walker, eds., The Definition of Morality, London: Methuen, 1970.
- 'Two Concepts of Morality', *ibid.*
- Darwall, S., Impartial Reason, Ithaca, New York: Cornell University Press, 1983.
- Davidson, D., 'How is Weakness of the Will Possible?', in Essays on Actions and Events, Oxford: Oxford University Press, 1980. First published in J. Feinberg, ed., Moral Concepts, Oxford: Oxford University Press, 1969.
- Elster, J., Ulysses and The Sirens, Cambridge: Cambridge University Press, 1979.
- Sour Grapes, Cambridge: Cambridge University Press, 1983.
- Foot, P., 'Moral Arguments', in her Virtues and Vices, Oxford: Blackwell, 1978. First published in Mind 67 (1958).

- 'Moral Beliefs', Virtues and Vices. First published in Proceedings of the Aristotelian Society, 59 (1958-9).
- 'Reasons for action and desires', *ibid.* First published in Proceedings of the Aristotelian Society, Suppl. Vol., 1972.
- 'Morality as a System of Hypothetical Imperatives', *ibid.* First published in The Philosophical Review, 81 (1972).
- 'A Reply to Professor Frankena', *ibid.* First published in Philosophy, 50 (1975).
- 'Are Moral Considerations Overriding?', *ibid.*
- 'Morality, Action and Outcome', in Ted Honderich, ed., Morality and Objectivity, London: Routledge & Kegan Paul, 1985.
- 'Utilitarianism and the Virtues' Mind, 94 (1985).
- Frankena, W., 'Obligation and Motivation in Recent Moral Philosophy', in W. Sellars and J. Hospers, eds., Readings in Ethical Theory, Second edition, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970. First published in A.I. Melden, ed., Essays in Moral Philosophy, Seattle: University of Washington Press, 1958.
- 'The Concept of Morality', in G. Wallace and A.D.M. Walker, eds., The Definition of Morality, London: Methuen, 1970. First published in University of Colorado Studies, Studies in Philosophy 3 (1967).
- Ethics, Second edition, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973.
- Thinking about Morality, Ann Arbor: University of Michigan Press, 1980.
- Gauthier, D., 'Thomas Hobbes: Moral Theorist', Journal of Philosophy 76 (1979).
- 'Deterrence, Maximization, and Rationality', Ethics, 94 (1984).
- Glover, J., Causing Death and Saving Lives, Harmondsworth: Penguin Books, 1977.
- What Sort of People Should There Be?, Harmondsworth: Penguin Books, 1984.
- Grice, G.R., The Grounds of Moral Judgement, Cambridge: Cambridge University Press, 1967.
- Griffin, J., Welfare, forthcoming.
- Hare, R.M., 'Universalisability', reprinted in his Essays on the Moral Concepts, London: Macmillan, 1972. Originally published in the Proceedings of the Aristotelian Society 55 (1954-5).
- Freedom and Reason, Oxford: Oxford University Press, 1963.
- 'Pain and Evil', in Feinberg ed., Moral Concepts, Oxford: Oxford University Press, 1969. First published in Proceedings of the Aristotelian Society, Suppl. Vol. 38 (1964).
- 'Wrongness and Harm', Essays on the Moral Concepts.
- Moral Thinking, Oxford: Oxford University Press, 1981.
- Harman, G., 'Moral Relativism Defended', Philosophical Review, 84 (1975)
- The Nature of Morality, New York: Oxford University Press, 1977.

- 'Relativistic Ethics: Morality as Politics', in P. French, T. Uehling Jr., H. Wettstein, eds., Midwest Studies in Philosophy, Vol. III: Studies in Ethical Theory, (1978).
- 'Justice and Moral Bargaining', Social Philosophy and Policy, 1 (1983),
- 'Is There a Single True Morality', in D. Copp and D. Zimmerman, eds., Morality, Reason and Truth, Totowa, New Jersey: Rowman & Allanheld, 1985.
- Hart, H.L.A., The Concept of Law, Oxford: Oxford University Press, 1961.
- Hobbes, T., Leviathan, 1651.
- Hume, D., A Treatise of Human Nature, Second edition, L.A. Selby-Bigge, ed., Oxford: Oxford University Press, 1978. First published in 1739-40.
- An Enquiry Concerning the Principles of Morals, Third edition, L.A. Selby-Bigge, ed., Oxford: Oxford University Press, 1975. First published in 1751.
- Kagan, S., 'Does Consequentialism Demand Too Much? Recent Work on the Limits of Obligation', Philosophy and Public Affairs, 13 (1984).
- Kavka, G., 'The Reconciliation Project', in D. Copp and D. Zimmerman, eds., Morality, Reason and Truth, Totowa, New Jersey, Rowman & Allanheld, 1985.
- Lovibond, S., Realism and Imagination in Ethics, Oxford: Blackwell, 1983.
- Lukes, S., 'Taking Morality Seriously', in T. Honderich, ed., Morality and Objectivity, London: Routledge & Kegan Paul, 1985.
- Lyons, D., Ethics and the Rule of Law, Cambridge: Cambridge University Press, 1984.
- Mackie, J.L., 'Sidgwick's Pessimism', Persons and Values, vol. II of his Selected Papers, Oxford: Oxford University Press, 1985. First published in Philosophical Quarterly 26 (1976).
- Ethics: Inventing Right and Wrong, Harmondsworth: Penguin Books, 1977.
- 'The Law of the Jungle: Moral Alternatives and Principles of Evolution', Persons and Values. First published in Philosophy, 53 (1978).
- Hume's Moral Theory, London: Routledge & Kegan Paul, 1980.
- 'Co-operation, Competition, and Moral Philosophy', Persons and Values. Originally published in A.M. Colman, ed., Cooperation and Competition in Humans and Animals, Wokingham: Reinhold, 1982.
- 'Morality and the Retributive Emotions', Persons and Values. First published in T. Stroup, ed., Edward Westermarck: Essays on His Life and Work, Helsinki: Acta Philosophica Fennica, 1982.
- 'Rights, Utility, and Universalization', Persons and Values. First published in R. Frey, ed., Utility and Rights, Minneapolis: University of Minnesota Press, 1984.

- 'Norms and Dilemmas', Persons and Values.
- 'The Three Stages of Universalization', Persons and Values.
- McDowell, J., 'The Role of Eudaimonia in Aristotle's Ethics', in A. Rorty, ed., Essays on Aristotle's Ethics, London: University of California Press, 1980.
- 'Non-Cognitivism and Rule-Following', in S. Holtzman and C. Leich, eds., Wittgenstein: To Follow a Rule, London: Routledge & Kegan Paul, 1981.
- 'Are Moral Requirements Hypothetical Imperatives?', Proceedings of the Aristotelian Society, suppl. vol. 52 (1978).
- Midgley, M., 'The Objection to Systematic Humbug', in her Heart and Mind, Brighton: Harvester Press, 1981. Originally published in Philosophy 53 (1978).
- Mill, J.S., Autobiography, J. Stillinger, ed., Boston: Houghton Mifflin Co., 1964. First published in 1874.
- Nagel, T., Possibility of Altruism, Princeton: Princeton University Press, 1970.
- 'Subjective and Objective', in his Mortal Questions, Cambridge: Cambridge University Press, 1979.
- 'The Limits of Objectivity', in S. McMurrin, ed., The Tanner Lectures on Human Values, vol. I, Cambridge: Cambridge University Press, 1980.
- 'The Excessive Demands of Impartiality', London Review of Books, 1-14 July 1982.
- Nielsen, K., 'Why Should I Be Moral?', in W. Sellars and J. Hospers, eds., Readings in Ethical Theory, Second edition, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.
- Norman, R., Reasons for Action, Oxford: Blackwell, 1971.
- Nowell-Smith, P., Ethics, Oxford: Blackwell, 1957.
- Overvold, M., 'Self-Interest and the Concept of Self-Sacrifice', Canadian Journal of Philosophy, 10 (1980)
- 'Self-Interest and Getting What You Want', in H. Miller and W. Williams, eds., The Limits of Utilitarianism, Minneapolis: University of Minnesota Press, 1982.
- Parfit, D., Reasons and Persons, Oxford: Oxford University Press, 1984.
- Plato, The Republic.
- Prichard, H.A., 'Does Moral Philosophy Rest on a Mistake', in W. Sellars and J. Hospers, eds., Reading in Ethical Theory, Second edition, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970. First published in Mind 21 (1912).
- Quinton, A., Utilitarian Ethics, London: Macmillan, 1973.
- Railton, P., 'Alienation, Consequentialism, and the Demands of Morality', Philosophy & Public Affairs, 13 (1984).
- Rawls, J., A Theory of Justice, Cambridge, Mass.: Harvard University Press, 1971.
- 'Kantian Constructivism in Moral Theory', Journal of Philosophy, 77 (1980).
- Raz, J., 'Introduction', in Raz, ed., Practical Reasoning, Oxford: Oxford University Press, 1978.

- Richards, D.A.J., A Theory of Reasons for Action, Oxford: Oxford University Press, 1971.
- Ryle, G., 'Pleasure', in J. Feinberg, ed., Moral Concepts, Oxford: Oxford University Press, 1969. Originally appeared in Proceedings of the Aristotelian Society, Supp. vol. 28 (1954).
- Scanlon, T., 'Rights, Goals, and Fairness' in S. Hampshire, ed., Public and Private Morality, Cambridge: Cambridge University Press, 1978.
- Scheffler, S., The Rejection of Consequentialism, Oxford: Oxford University Press, 1982.
- Schelling, T., The Strategy of Conflict, Cambridge, Mass.: Harvard University Press, 1960.
- Seabright, P., 'Character', London Review of Books, 5 September 1985.
- Sen, A.K., 'Rights and Agency', Philosophy and Public Affairs, 11 (1982).
- 'Rational Fools', in F. Hahn and M. Hollis, eds., Philosophy and Economic Theory, Oxford: Oxford University Press, 1979. First published in H. Harris, ed., Scientific Models and Man. The Herbert Spencer Lectures 1976, Oxford: Oxford University Press, 1979.
- Sidgwick, H., Methods of Ethics, Seventh edition, London: Macmillan, 1907. Reissued by Hackett Publishing Company, Indianapolis, 1981.
- Singer, P., Practical Ethics, Cambridge: Cambridge University Press, 1979.
- Slote, M., Common-sense Morality and Consequentialism, London: Routledge & Kegan Paul, 1985.
- Goods and Virtues, Oxford: Oxford University Press, 1983.
- Strawson, P.F., 'Social Morality and Individual Ideal' Freedom and Resentment and Other Essays, London: Methuen, 1974. First published in Philosophy, 36 (1961).
- Stroud, B., Hume, London: Routledge and Kegan Paul, 1977.
- Toulmin, S., An Examination of the Place of Reason in Ethics, Cambridge University Press, 1950.
- Urmson, J.O., 'The Goals of Action', in A. Goldman and J. Kim, eds., Values and Morals, Dordrecht, Holland: D. Reidel, 1978.
- Warnock, G.J., Contemporary Moral Philosophy, London: Methuen, 1967.
- Winch, P., 'The universalizability of moral judgments', The Monist 49 (1965).
- Williams, B., Ethics and the Limits of Philosophy, Cambridge, Mass.: Harvard University Press, 1985.
- Morality: an Introduction to Ethics, New York: Harper & Row, 1972. Reissued by Cambridge University Press, 1976.
- 'A Critique of Utilitarianism', in J.J.C. Smart and B. Williams, Utilitarianism: For and Against, Cambridge: Cambridge University Press, 1973.
- 'Ethical Consistency', his Problems of the Self, Cambridge: Cambridge University Press, 1973.

- 'Conflicts of Values', in his Moral Luck, Cambridge: Cambridge University Press, 1981.
- 'Internal and External Reasons', Moral Luck.