The motivational capacity of moral theories

Karen Elizabeth van den Broek

A thesis submitted for the degree of
Doctor of Philosophy of
The Australian National University

January 2001
Statement

This is to certify that, unless otherwise indicated, this thesis is entirely my own work. It is the result of research carried out by me while a candidate for the degree of Doctor of Philosophy, in the Philosophy Program, Research School of Social Sciences, The Australian National University.

Karen Elizabeth van den Broek
In this thesis I argue that normative moral theorising cannot proceed independently of considerations of how people could be motivated to act. A justificatory moral reason must plausibly be able to be someone’s motivational reason. Whether or not a moral theory is able to give plausible motivational reasons is a test of the adequacy of the theory. This idea I refer to as the ‘motivational test’ of moral theories. I put three moral theories to this test, posing the question, ‘could this theory plausibly motivate moral action?’ The three theories are: consequentialism, Kantian ethics and the ethic of care. I argue that none of these theories can plausibly provide a complete set of motivating reasons for action. Although they are able to motivate moral action some of the time, they cannot motivate it all of the time. Consequentialism, understood as requiring the promotion of agent-neutral value, is limited because it cannot motivate acts of friendship or respect for persons for their own sakes. Kantian ethics, while being able to motivate respect for persons, cannot motivate friendship and is too narrow in its motivational scope. The ethic of care is able to motivate friendship but cannot motivate action requiring an impartialist outlook. Each of the three theories contains motivational gaps which are able to be covered by one of the other two theories. Thus, while each of the theories can only provide an incomplete set of moral motivational reasons, collectively it may well be that these three theories can provide a complete set of such reasons. I take this to be an argument for moral pluralism.
Dedicated to the memory of my mother.
## Summary table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>xv</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>xix</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1 Internal and external reasons</td>
<td>13</td>
</tr>
<tr>
<td>2 Reasons based on consequences</td>
<td>55</td>
</tr>
<tr>
<td>3 Reasons based on duty</td>
<td>101</td>
</tr>
<tr>
<td>4 Reasons based on care</td>
<td>149</td>
</tr>
<tr>
<td>5 A plurality of reasons</td>
<td>169</td>
</tr>
<tr>
<td>Conclusion</td>
<td>207</td>
</tr>
<tr>
<td>Bibliography</td>
<td>211</td>
</tr>
</tbody>
</table>
## Contents

**Preface**  
**Abbreviations**  
**Introduction**  

### 1 Internal and external reasons

1.1 Williams’s distinction  
1.2 In defence of external reasons: seeing matters aright  
1.2.1 The idea of conversion  
1.2.2 Objections  
(i) Having an internal reason but not being aware of it  
(ii) Having an external reason or having no reason?  
1.3 In defence of external reasons: the justificatory standpoint  
1.3.1 The role of external reasons  
1.3.2 Objection  
1.4 The motivational test  
1.4.1 The relationship between internal and external reasons  
(i) External reasons as a limit to internal reasons  
(ii) Internal reasons as a limit to external reasons  
1.4.2 External reasons and moral theories  

**Conclusion**
## 2 Reasons based on consequences

2.1 Features of consequentialism
   2.1.1 Pettit’s consequentialism
   2.1.2 The plausibility of being motivated by consequences
   2.1.3 Agent-neutral and agent-relative values
   2.1.4 Agent-relative value and universalizability

2.2 Consequentialism and respect for others
   2.2.1 Indirect consequentialism
   2.2.2 Two kinds of indirect consequentialism
      (i) Indirect consequentialism and secondary motivation
      (ii) Indirect consequentialism and primary motivation
   2.2.3 Respect as an agent-relative value

2.3 Consequentialism and friendship
   2.3.1 Friendship as an agent-relative value
   2.3.2 Indirect consequentialism and friendship

2.4 Two consequentialist replies
   2.4.1 Truth-conditions versus acceptance-conditions
      (i) Moderate objective consequentialism
      (ii) Extreme objective consequentialism
   2.4.2 Kupperman’s argument

Conclusion

## 3 Reasons based on duty

3.1 Acting from duty, Kantian-style
   3.1.1 Doing one’s duty
   3.1.2 Duty as a plausible motive

3.2 Duty and friendship
   3.2.1 The importance of affection in friendship
   3.2.2 The demands of friendship

3.3 Duty and affection as coexistent
3.4 Friendship as outside the scope of morality  
3.4.1 Moral features of friendship  
3.4.2 The calibre of Kantian friendship  
  (i) The emotional life of the Kantian friend  
  (ii) The detached stance of the Kantian friend  
3.5 Liars and murderers: the narrowness of Kantian ethics  

Conclusion  

4 Reasons based on care  
4.1 Features of an ethic of care  
  4.1.1 Caring as an ethical activity  
  4.1.2 Care as a plausible motive  
4.2 The limits of care as a motive  
  4.2.1 The narrow ambit of care  
    (i) The attribution of moral status  
    (ii) The absence of social criticism  
  4.2.2 The ethic of care as a professional ethic  

Conclusion  

5 A plurality of reasons  
5.1 An argument for pluralism  
5.2 Problems for pluralism  
  5.2.1 The problem of simplicity  
  5.2.2 The problem of coherence  
  5.2.3 The problem of conflicts  
    (i) Conflicting principles  
    (ii) The importance of judgement
5.3 Pluralism and the virtues 183

5.3.1 Virtue ethics and the three theories 184

5.3.2 Features of virtue ethics 189

(i) Agent-centredness 190

(ii) The inner moral life 191

(iii) The shape of a life 197

(iv) Moral guidance 198

(v) Self-regard 199

5.3.3 A pluralistic theory of the virtues 200

Conclusion 204

Conclusion 207

Bibliography 211
Preface

It seems proper, by way of prelude, to make some mention of the reasons behind the chosen topic. When I first began reading the classic texts in moral philosophy – from Aristotle to Kant to contemporary writers – I found myself repeatedly struck by two discordant thoughts. On the one hand was the thought, ‘how true,’ or ‘how fine.’ On the other hand was the thought, ‘but this will not work if it is pushed to its limits.’ And so it was. When reading Kant I became a Kantian; when reading Mill, a utilitarian; when feminism, a feminist. But never completely, never unreservedly. Underpinning this twofold reaction was a vaguely Socratic notion that ethics had to do with how one should live. One could live as each moral theory prescribed, I thought, but only up to a point.

When faced with such chameleon indecision, one hopes that, with more reflection, more conversation, and with more experience in the ways of the world, one or other dominant theory will eventually emerge as the right one. ‘I am Young and Foolish,’ I thought. ‘Time will cure all.’

Alas, it was not to be. The ambivalence towards various approaches to ethics has not resolved itself. On the contrary, it has intensified. For we are complex creatures who are capable of – and find ourselves in situations calling for – a variety of ethical responses. Banal though such a statement may be, I find there is no escaping it. Acting with a view to bringing about the best consequences is right some of the time, but not all of it; acting from respect for the autonomy of others is important, but it is not the only consideration; acting for concern for close attachments has its place, but only a place. And so it is. Here I stand, a reluctant pluralist. I am driven to that conclusion only by the conviction that attempts to prescribe ethical conduct in terms of a single principle of conduct, or a single kind of value, are unsuccessful. I say
‘reluctant,’ since it occurs to me that pluralism lacks the coherence, definiteness and simplicity of other approaches. In the thesis, I argue against three approaches to ethics to a conclusion of pluralism.

In formulating my thoughts, I was influenced by several essays. Michael Stocker’s ‘The schizophrenia of modern ethical theories’ was a paper to which I kept returning. The problem with modern ethical theories, says Stocker, is that they fail to motivate. According to Stocker, if a theory cannot be lived by, it is a bad theory. This thought became an underlying theme of the thesis. Much of Bernard Williams’s work also has an affinity with the idea that it is a mistake for moral theorising to proceed independently of what it is plausible for people to be motivated by. His short but provocative paper ‘Internal and external reasons’ is a variation on this theme. Although I differ from Williams in the detail, I proceed throughout the thesis in a spirit responsive and sympathetic to his contribution to moral philosophy. Finally, Max Weber’s splendid essay, ‘The profession and vocation of politics,’ gave me a framework for understanding that morality can, and often does, place irreconcilable demands upon us. Although his essay was originally written as a lecture for graduating students who might be contemplating a career in politics, it was clear that Weber’s essay had lessons for everyone, not just politicians. From Stocker and Williams, then, I learnt to be sceptical of how far moral theorising could plausibly go, and from Weber I learnt to trust my pluralistic intuitions.

I am fortunate indeed to have had the opportunity to serve my apprenticeship in the stimulating intellectual environment provided by the Philosophy Program in the Research School of Social Sciences. I would like to thank all the staff, students and visitors, too many to mention by name, with whom I had contact in the time I was there between 1995 and 2000. I am especially grateful to Michael Smith who, as my supervisor, gave me numerous useful suggestions for improving my work. From him, and from the many eminent philosophers who flanked the corridors of the Program, I learnt the importance of striving for clarity and precision. I would also like to thank those who read and commented on earlier drafts of my thesis, or parts thereof, including Philip Gerrans, Andrew Gleeson, Karen Jones, Thomas Mautner and Peter Wylie. I thank Susan Tridgell in particular, for her relentless encouragement, and Sarah Bachelard, Hamish Cowan and Barbara Nunn for their philosophical conversation.

1 Stocker, 1987. See also the editors’ introduction to the paper, p. 36.
2 Williams, 1981d.
An earlier version of Chapter four was published as a chapter in *Health care, law and ethics* in 1997, and I would like to thank the editor of that book for permission to use the paper again here. 4

In writing the thesis I faced a practical problem which cannot pass unmentioned since the solution of it has placed upon me a large debt of gratitude to so many people. Due to a long-standing disability, I was—and still am—unable to handwrite, type or even use speech-recognition software. This would have posed an insurmountable obstacle had it not been for the support of Margaret Miller and Trevor Allan of the University’s Disability Office. The understanding they showed me went beyond the call of duty, and I thank them for it. I thank them too for enlisting a professional typist, Robyn Swadling, to type to my dictation. The entire thesis was dictated in short bursts to Robyn, my chief companion-in-writing since 1997, and a group of volunteers. I thank Robyn for her goodwill and sheer stamina.

Mastering this unusual method of writing a PhD dissertation was not without challenge. My inspiration was the generosity and forbearance of the many people who, at one time or another, volunteered their time to assist me. Foremost among them are Linda Gosnell, Roy Jeeves, Gabrielle McCudden, Ria van de Zandt and Trish Worth. Warm thanks go to Lynda Hilton, whose weekends are now her own again. I am much indebted to Glenda Davis who, like Dorothea to Mr Casaubon in *Middlemarch*, hounded me to make a start, and who was there at the end to help with proof-reading. As for Betty Kat, it is hard to do justice to the commitment she has shown to this project. For years she arrived at my door at the appointed hour, always good-humoured, always willing, always full of keen advice on points of scholarship and expression. It is now my pleasure to share with her, and all the others who made a contribution, the victory of completion.

I owe much to Mark Pallas who, when the day was done, was ever willing to lend me his sympathetic ear. I thank him for gifts spoken and unspoken.

Finally, I owe an enduring debt to Patrick Brady, who first showed me the path of philosophy and urged me on its way.  

---

### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Title</th>
<th>Edition Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRL</td>
<td><em>On a supposed right to lie from philanthropy</em>. First published 1797. Pagination is from volume 8 of the Berlin Academy Edition.</td>
<td></td>
</tr>
</tbody>
</table>


Introduction

Procrustes, you will remember, stretched or chopped down his guests to fit the bed he had constructed. But perhaps you have not heard the rest of the story. He measured them up before they left next morning, and wrote a learned paper 'On the Uniformity of Stature of Travellers' for the Anthropological Society of Attica.

Sir Arthur Eddington
The Tarner lectures, Cambridge, 1938

Let us begin at the end. By stating the conclusion of the thesis at the very outset we can become clear about where we are headed. Everything that is contained in the forthcoming five chapters is designed to make moral pluralism seem plausible.

While some writers assume pluralism, taking it for granted, I hope to conclude it after detailed examination of three moral theories which are not pluralistic. One way of arguing for pluralism, as opposed to merely asserting it, is to consider what our experience of practical deliberation reveals. This means looking to our experience of, say, dilemmas or akrasia. Here is how the argument from dilemmas works. Dilemmas, it is claimed, are those situations where two or more options are equally compelling; one has an overriding reason to do both A and not-A. Whichever choice one eventually makes there may be a moral remainder, or what Bernard Williams calls 'regret' for what was

\[1\] Thomas Nagel, for example, asserts that there are five types of value. (Nagel, 1979c, pp. 129-130.) Charles Larmore claims that there are three principles of practical reason. (Larmore, 1987, Ch. 6.)

\[2\] Stocker, 1997, p. 199.

\[3\] Walter Sinnott-Armstrong describes dilemmas in terms of overriding reasons. (Sinnott-Armstrong, 1985, p. 321.)
missed."\textsuperscript{4} That is to say, in such situations one feels regret or remorse about the option one did not take. This regret, it is claimed, points to there being a plurality of values.\textsuperscript{5} One has regret in such situations because one is unable to respond to value as one should.

There is an important assumption underlying this train of thought: our common experience cannot be dismissed as morally irrelevant. The pluralist insists that the emotional aftermath of dilemmatic situations cannot be disregarded, that it is central to our understanding of this kind of conflict. To understand the nature of dilemmas we need to look to how people experience and respond to them. The idea is that the moral life, in all its ‘complexity and vastness,’ must underpin our theoretical understanding of morality.\textsuperscript{6} This is a distinguishing feature of pluralistic approaches.

On the other hand, the view that there are no such things as dilemmas is supported by principles of deontic logic, such as the ‘ought implies can’ principle.\textsuperscript{7} According to this principle, if I ought to do action \textit{A} then I can do it. Similarly, if I ought to do action \textit{B} then I can do it. Therefore, if I ought to do \textit{A} and \textit{B} then I \textit{can} do \textit{A} and \textit{B}. But a dilemma is putatively a situation in which I ought to do both \textit{A} and \textit{B}, but I \textit{cannot} do both. Therefore, goes the argument, dilemmas are a logical impossibility: there can only ever be one option that places an overriding obligation on us. This conclusion suits both the consequentialist and the Kantian deontologist.\textsuperscript{8}

It is open to the pluralist who admits the existence of dilemmas, however, to challenge the ‘ought implies can’ principle. The principle is not infallible, it might be said; the dilemma is the counter-example to it. According to this line of thought, what we know by experience cannot be relegated to the philosophical dustbin by principles of logic or other pre-determined theoretical standpoints.

The argument for pluralism presented in this thesis also proceeds by appealing to common experience and understanding. I do not, however, appeal specifically to dilemmas. Rather, I proceed from the claim that one cannot engage in theoretical speculation about normative ethical matters independently

\textsuperscript{4} Williams, 1973b, p. 170.
\textsuperscript{5} Stocker (1997) makes this argument.
\textsuperscript{7} For consequentialist views see, for example, Mill, 1987 and Norcross, 1995. (Mill’s work first published 1843.) Kant defines moral duty in such a way as to rule out conflicts of duty (DR, 224). I take the liberty here of referring to Kantian ethics as ‘deontology,’ but acknowledge that whether Kant’s ethics can properly be described in this way is a matter of debate. (See Herman, 1993g.)
of considerations of what it would mean for someone to hold the views in question and act upon them. That is to say, a moral theory cannot be examined, developed or upheld independently of considerations of what kind of life would result for someone who took the theory seriously and exemplified it. If an outlook, decision or action is right, according to some ethical theory, it must plausibly be able to be the outlook, decision or action of some actual person, somewhere.9

But what would it mean to ‘exemplify’ a theory or to live according to it? One way of answering this question would be to say that a moral theory must be able to provide plausible motivating reasons for action. This is the interpretation for which I shall argue. I do not claim to be the first to suggest that moral theories must be able to motivate. The separation of the justificatory role of moral theories from their motivational plausibility has met with the disapproval of Michael Stocker, for one. ‘Modern ethical theories,’ complains Stocker, ‘deal only with reasons, with values, with what justifies. They fail to examine motives and the motivational structures and constraints of ethical life.’10 The subsequent split, he says, between one’s motives, on the one hand, and one’s reasons or justifications, on the other hand, may properly be called ‘moral schizophrenia.’ Let us be clear about this schizophrenia. Suppose we accept the contention that a moral theory tells us the reason, \( R \), why an action or a certain class of actions, \( A \), is right. The problem arises when \( R \) is then unable to motivate us to do \( A \). We believe the theory to be true but we are unable to act upon it because it presupposes an implausible moral psychology. Stocker claims that most modern ethical theories suffer this malady. His main targets are consequentialism and Kantian deontology which, he says, cannot allow for ‘love, friendship, affection, fellow feeling, and community.’11

It is useful to spell out where the difference lies between Stocker and me regarding the importance of ethical theories being able to motivate. In this thesis I examine three ethical theories and show that none of them is able to provide a plausible set of motivating reasons. Instead of concluding, as Stocker does, that these theories therefore require a schizophrenic response on the part of those who subscribe to them, I conclude that each of the theories is incomplete. For while none of them is able to provide a complete set of moral motivations, they are each able to provide some motivations. That is, they can plausibly motivate some of the time, but not all of the time. Additionally, while

---

9 See Hollis, 1995, especially p. 177.
10 Stocker, 1987, p. 36.
11 Stocker, 1987, p. 40. Egoists have a similar problem. (Stocker, 1987, p. 38.)
they are individually limited, they may well be collectively complete. This is the argument for moral pluralism.

While Stocker has explicitly articulated that a theory can be criticised on the grounds of its inability to motivate, there are others, it seems to me, whose criticisms of various theories assume this point. I have in mind two examples of this assumption, one from Williams, the other from Susan Wolf.

Does not Williams's seminal criticism of utilitarianism presuppose that utilitarianism should be able to motivate a person to action? Recall, briefly, his frequently cited example of Jim and Pedro. Jim, the intrepid explorer, encounters Pedro, the tyrant, in the heart of the jungle. Pedro is about to kill twenty Indians, but in honour of Jim’s arrival, puts it to Jim that the Indians shall be set free if Jim will shoot just one of them. The question, then, is whether Jim should shoot one to save the nineteen, or refuse to shoot any, and have Pedro shoot all twenty. Williams says that the decision is clear-cut for the consequentialist: Jim should shoot the one because one dead is a better state of affairs than twenty dead. Williams protests that the decision is not so simple. There may be more to take into account than just the numbers of dead bodies which will result from the two actions. It may be relevant to Jim’s decision, not just that one person is dead rather than twenty, but that the one dead was shot by him, Jim, by his hand. Killing an innocent man may be something Jim will refuse to do, whatever the consequences.

The point here is not whether Jim should or should not shoot the Indian. Rather, the interesting issue is Williams’s strategy in pointing to deliberations at all. In showing how limited utilitarianism is, he draws our attention to possible motivating factors other than consequentialist ones. Williams draws our attention to the question of what it would mean for a person to go through life always making decisions based upon utility outcomes. He concludes that consequentialist reasoning, if taken seriously, would render a life unlivable: one would need to view one’s projects impartially, as part of an overall utility calculus. Because of this, he says, consequentialist reasoning amounts to an attack on one’s integrity. By saying that consequentialism is too

---

12 Williams, 1973a.
13 Nor is it Williams’s point. The story of Jim and Pedro is not to be taken as an example of a utilitarian choice being the wrong one. ‘If that had been the point,’ Williams later remarked, ‘I would have had to tell you more about the cases than I did.’ The point is that utilitarianism gives a depleted way of understanding such cases. There is much more to be said about such situations than utilitarianism allows for, whatever choice is eventually made. (Williams, 1994, pp. 11-12.) Nagel’s paper, ‘War and massacre,’ however, does give a defence of the non-consequentialist option in such situations. (Nagel, 1979b.)
demanding because it does not permit us to place special weight on personal projects and commitments, it is meant that consequentialism would be too demanding on the life of any particular agent to understand ethics in this way. The assumption here is that, in deciding whether a particular theoretical stance is right, one needs to consider what it would be like for an agent to live according to that stance, where ‘living according to’ means, at the very least, ‘being motivated by.’

Nor does Williams’s strategy meet with uniform condemnation from those who would defend consequentialism. Frank Jackson, in responding to Williams’s argument, does not deny that consequentialism must be able to give motivating reasons. Rather, he argues, in effect, that consequentialism is able to accommodate the kinds of motives that Williams has said it cannot. ‘Consequentialism,’ says Jackson, ‘can make plausible sense of the moral agent having and giving expression in action to a special place for family, friends, colleagues, chosen projects, and so on and so forth.’ He sees it as essential for consequentialism that the connection between doing as the theory enjoins on the one hand, and having a meaningful life on the other, be sustained. It would be ‘chilling,’ he says, to break this connection.

The other example is to be found in Wolf’s paper, ‘Moral saints.’ Wolf argues here that there are limits to what morality can plausibly and desirably demand of us. She asks us to consider what sort of person a moral saint would be, then puts it to us that such a person would have a life which was ‘strangely barren.’ Such saints would have, and would cultivate, the qualities which would enable them to treat others as justly and kindly as possible, Wolf says. They would be so preoccupied by being as morally good as possible that they would neglect the non-moral interests and personal qualities which we think make a ‘healthy, well-rounded, richly developed character.’ Wolf is glad that neither she nor those she cares about are saints, on account of this attendant narrowness. Wolf’s target here is not the excessive

16 Jackson, 1991, p. 461. He argues a variant of John Austin’s idea that ‘we act to produce the most good when the sphere wherein we act is the most restricted and the most familiar to us.’ (Austin, 1996, Lecture 4, p. 163. First published 1832.) Jackson has in mind a form of indirect consequentialism. I consider indirect consequentialism in detail in § 2.2 and § 2.3.
17 Wolf, 1982.
19 Wolf, 1982, pp. 419, 420. Harry Frankfurt also takes the view that morality is not everything. ‘Even people who care a great deal about morality generally care still more about other things,’ he says. (Frankfurt, 1988, p. 81.)
Wolfs strategy in this paper is to examine the kinds of motives which an individual could desirably have. It turns out, she says, that they consist in both moral and non-moral considerations. The conclusion? A warning for moral theory and theorists. Moral theorists cannot ignore the extent to which their theories emphasise the moral over the non-moral. Again, what is interesting about Wolf’s paper is not whether she is right or wrong in taking sainthood to be so unappealing. The point, rather, is that she has assumed that it is the task of moral theories to provide plausible motivating reasons.20

The ability of a theory to give sound and plausible motivating reasons amounts to a test of the adequacy of a theory. Let us call this test the motivational test of moral theories. We have already seen that the idea behind the motivational test – that ethical theorising must not proceed independently of considerations of what it would amount to for an individual to live according to those theories – is present, both implicitly and explicitly, in the contemporary literature.

Having referred, in the broadest of terms, to the need for moral theories to provide plausible motivating reasons, three things now need to be done. The first is to make precise what would count as ‘providing plausible motivating reasons.’ That is to say, the motivational test must be better explained and justified. The second task is then to show whether various theories are capable of passing it. The third and final task is to point out what conclusions may be drawn from various theories either passing or failing the test. Accordingly, the thesis argues three claims.

(1) A necessary test of the adequacy of a moral theory is that the theory be able to provide a set of reasons which could plausibly motivate.
(2) Individually, none of the three moral theories that I nominate can provide such a set of reasons, but collectively it seems that they can.
(3) Therefore, we have an argument for moral pluralism.

These three claims provide the structure of the thesis. There are five chapters. In the first Chapter I argue for (1). I set up the motivational test, explaining what it means and how it is to be applied. In the three middle chapters I argue for (2). I apply the motivational test to three different moral

---

20 There are, of course, those who say she has misrepresented moral sainthood. See, for example, Baron, 1997, pp. 20-21.
theories, a chapter for each theory. I hope to show, through hypothesis and imaginative example, that this is a very tough test indeed for a theory to meet. I argue that none of the theories meets it. In the fifth and final chapter I argue for (3). I justify my selection of the three candidate theories and argue that, although none of the theories alone can provide a plausible motivating set of moral reasons, collectively they may well be able to. Even if not, we are in any case left with a conclusion of moral pluralism.

Let us be clear about the relative significance of these three tasks. The focus of the thesis is not pluralism itself. The pluralism is merely the end point of a certain argument. Similarly, the focus of the thesis is not the nature of moral motivation. The motivational test is simply a device for connecting moral theory to the moral life. It provides a much needed link between moral theory and moral practice. The bulk of the thesis is concerned with showing the limits of three dominant ethical theories. The same question is posed of each theory, namely, ‘What would follow if someone were to take the theory so seriously that she would be motivated to act according to it throughout her life?’ We want to send someone out into the world with only one particular moral theory for guidance. What kind of life – moral or otherwise – would he or she be able to have?

More needs to be said about each chapter. The task of Chapter 1 is to make the motivational test look plausible. A paper by Williams, ‘Internal and external reasons,’ provides a starting point for talking about the connection between moral theories, reasons and motivations. We have seen that in Utilitarianism: for and against, Williams is sceptical of utilitarianism being able to give an account of rightful action which is independent of the particular commitments of the person who has to take the action. In facing Pedro’s ultimatum, Jim may well have more to consider than the simple numbers of the dead which result from either option. Jim has to deliberate from his own particular standpoint, which may well include a conviction that killing innocent people is wrong. Similarly, in ‘Internal and external reasons,’ Williams expresses scepticism of the idea that people can have any reasons for acting at all which are independent of the projects, commitments and other motivations which give their lives structure and meaning.

Williams argues that an internal reason is one which satisfies an element in an individual’s motivational set, S. An external reason, he says, is one which has no such relationship to S. He contends that there are, in fact, only internal reasons. It only makes sense to talk of reasons, in other words, in the context

---

21 Williams, 1981d.
of what motivates a particular person (or, as it turns out, what would motivate her if she deliberated in the right way).

In the first chapter I aim to show that this is not quite right, that there are external reasons but they should not be understood as being independent of internal reasons. External (moral) reasons exist independently of any particular motivational set, and are themselves justified, I suggest, by moral theories. These external reasons perform a justificatory role. I argue that these two kinds of reasons – the (internal) motivational reasons and (external) justificatory reasons – are not, however, independent of each other: what counts as a good internal reason is limited by external reasons, and what can count as an external reason is limited by what it is plausible for a person to hold as an internal reason. This latter claim, that justificatory reasons must be able to be someone’s motivational reasons, provides the basis for the motivational test.

In the second, third and fourth chapters I examine three different theories respectively: consequentialism, Kantian deontology and the ethic of care. The chapters on consequentialism and Kantian ethics are the longest and most complex of the five chapters of the thesis. This is not surprising. It is commonly observed that the various versions of consequentialism and Kantian deontology are the dominant ethical theories of our time. It is no small task to show the ways in which these two ethical approaches are limited. We need to consider the various replies and counter objections which defenders of these theories could raise against the motivational test and the way it is applied.

All three theories fail the test for essentially the same reason. They each provide too limited a set of motives to be plausible. I argue that each of the theories has motivational ‘blind spots,’ meaning that each rules out certain kinds of important motives. However, each motivational blind spot is able to be covered by one or other of the remaining two theories.

Here, in summary, are the arguments. If one were a consequentialist one would not have the motives necessary for showing non-instrumental respect to others, or for having friendships. Consequentialism, that is, rules out respect and friendship. If one were a Kantian, on the other hand, one would be able to show due respect to others but there would be other problems. Again, one would not be able to have the motives necessary for friendship, and one would be too narrow in the scope of one’s moral outlook. Finally, if one were motivated by an ethic of care, one would have the motives necessary for friendship but would be excluded from having the impartialist outlook provided

---

by the previous two theories. Each theory prescribes *some* desirable motives but no theory has the *complete* set of desirable motivations.\(^{23}\)

Each theory gives an account of moral action but can only do so at the expense of omitting certain classes of motives, which, if not themselves moral are at least desirable by any common measure of a good life. Each theory suffers from what may be called a *Procrustean fallacy*.\(^{24}\) Just as Procrustes trimmed the travellers to fit the bed, so too each of our three theories trims our common experience, understanding and appellation of morality or a good life.

This brings us to the last chapter of the thesis. Chapter 5 is an exploration of what follows from the failure of each of these theories to meet the motivational test. I argue that their collective failure amounts to pluralism. Also in Chapter 5 I justify the selection of these three moral theories to the exclusion of other theories like contractualism and virtue ethics.

Having indicated the lines along which the thesis argues, I should now mention what it does *not* argue. There are a couple of issues which impinge upon the topic but which, in the interests of brevity, I do not wish to enter into. The first of these is an argument about the scope and authority of morality. The second has to do with the nature of motivation.

I said earlier that each of the three theories rules out various important motives. We need to clarify the nature of those motives. One obvious question is whether the motives in question are moral or non-moral. There are two ways the motivational test could work. The first way would be to show that a person who did as the theory prescribed would be morally depleted. That is, they would have too narrow a set of moral motives for us to admire or aspire to. The second way would be to show that such a person would be depleted in some other, non-moral sense.

The motivational test focuses primarily on the *moral* motives which a theory rules out. But the question of whether the motives which a theory rules out are best understood as moral or non-moral cannot be entirely avoided. This issue is particularly important in the context of Kantian ethics. One argument against Kantian ethics is that it rules out the motives necessary for close personal attachments such as friendship.\(^{25}\) But we need to be clear about whether it is the *moral worth* of friendly motives which is supposedly being ruled out, or the motives themselves. This question forces one to take a position on the scope of morality. Should friendship be understood as a moral

---

\(^{23}\) What counts as a ‘desirable’ or ‘plausible’ motive is discussed in § 1.4.2.

\(^{24}\) See the epigram of the Introduction.

\(^{25}\) This argument is presented in Chapter 3, §§ 3.2-3.4. See n. 3 of that chapter for advocates of this kind of criticism.
good or a non-moral good? To circumvent a debate on this question I shall argue that friendship is ruled out as both a moral and a non-moral good. Even if we accept the view that friendship has no moral worth, we can agree that it is important. We can agree with Wolf in this case that a well-rounded character has both moral and non-moral motives and interests.26 I hope to make it plausible that friendship is best considered as a moral good, but this shall not be necessary to the argument against Kantian ethics. If a theory rules out friendship, then whether the friendship is understood as having moral worth or not, it will be an indictment of the theory, since, as Aristotle says, 'no-one would choose to live without friends even if he had all the other goods.'27

The second debate to be sidestepped relates to the nature of motivation. Insofar as it is possible we will put to one side the question of whether one should be an internalist or externalist about moral motivation. Internalism and externalism in this context refer to theses about the connection between moral judgement and motivation. The internalist thinks that moral judgements are necessarily connected to moral motivation. ‘To think that you ought to do something is to be motivated to do it. To think that it would be wrong to do something is to be motivated not to do it.’28 The externalist, however, holds that one can make a moral judgement without having any motivation in the direction of the judgement. The variants of each position – and their respective virtues and limitations – have been well documented.29 The connection between moral obligation and motivation does not require our direct attention. Rather, for our purposes the relevant question is, what would acting upon our obligations rule out in our lives? Similarly, we will not be focussing on the question of whether a Humean or Kantian account of motivation is the more plausible. We will not be able to dismiss this question altogether, however, since it necessarily arises in the context of Kantian ethics in the third chapter. But we need to bear in mind that the thesis is not so much concerned with the mechanisms of motivation as with the content of it. Or, to put it another way, meta-motivational questions will, for the most part, remain in the background, out of focus. Our interest lies in the what of motivation, rather than the how.

Many of the arguments in the forthcoming pages are familiar ones in the literature. For example, the idea that impartialist theories such as

---
26 Wolf, 1982.
27 NE, 1155a, 5.
29 For discussions of these positions see Audi, 1997, Ch. 1; Blackburn, 1998, pp. 60-66; Brink, 1989, Ch. 3; Dancy, 1993, Appendix 1; Darwall, 1983, Ch. 5; Falk, 1952; Frankena, 1958; Harman, 1977; Korsgaard, 1986a; McNaughton, 1988, § 2.3, Ch. 9; Smith, 1972; Smith, 1994; Svavarsdóttir, 1999.
consequentialism and Kantian ethics have difficulty accommodating close personal attachments is not new.\footnote{See Chapter 2, n. 72 for authors who criticise consequentialism along these lines. For similar criticisms of Kantian ethics see Chapter 3, n. 3.} This thesis makes a contribution to the existing body of literature in two ways. First, it adds to various conversations between consequentialists and their critics, Kantians and their critics and feminist care ethicists and their critics. Secondly, it gives a framework for bringing these conversations together, showing what they have in common and what conclusions we may draw from them.
1

Internal and external reasons

His reasons are as two grains of wheat hid in two bushels of chaff: you shall seek all day ere you find them; and when you have them, they are not worth the search.

Bassanio
The merchant of Venice

THERE is a philosophical question that just will not go away. Plato raised it in The Euthyphro: is something holy because the gods love it or do the gods love it because it is holy?

A modern version of the Euthyphro question refers to ‘a rational agent,’ rather than gods. The question then becomes: does a rational agent act from objectively good reasons, or are good reasons those which a rational agent would have? The issue has to do with what grounds reasons: is it the motives of the agent who has the reasons, or is it independent standards? Answers to this question divide along what may be called ‘Humean’ and ‘Kantian’ lines, with the Kantian keeping Socrates company on the latter side of the question.¹

The Kantian position is that, as Martin Hollis puts it, ‘Rational agents act from

¹ Korsgaard would dispute this reading of Kant. Korsgaard says that, in Kant’s view, we make something valuable by choosing it as an end. On her reading, Kant comes down on the subjectivist side of the Euthyphro question. But as Rae Langton points out, the laws of morality are not chosen; they have a practical necessity. Kant says that ‘it is not because the law interests us that it has validity for us ... instead, the law interests because it is valid for us.’ (Korsgaard, 1983, pp. 182 ff.; Langton, 2000 and G, 460-461.) Korsgaard and Langton are both right. Kant arguably takes the subjectivist side of the Euthyphro question when it comes to non-moral value, but takes the objectivist side in relation to moral value. But even as a subjectivistic, Kant does not think value is conferred through our desires but through our reasoned choices.
objectively good reasons, whose merits are conceptually independent of their current desires.\footnote{Hollis, 1987, p. 74.} The Humean position denies that there are objectively good reasons, and holds that reasons are always dependent on the current desires of the agents who have the reasons.

In this chapter I discuss a paper by Bernard Williams, ‘Internal and external reasons,’ wherein he argues a variant of the Humean position. Williams describes Humean-style reasons as internal reasons. One has an internal reason to φ, he says, if φ-ing furthers an element in one’s motivational set, \footnote{Williams, 1981d, p. 101.} \footnote{Hollis, 1995, p. 171.} External reasons, by contrast, are Kantian-style reasons, namely those which everyone has, whether or not everyone is motivated by them.

Williams denies that there are any such external reasons. He says that what might be thought of as external reasons, on close examination, dissolve into internal reasons. What is to be gained by denying external reasons? Williams’s target is the idea that there are free-floating reasons which compel all individuals, on pain of irrationality. His point is that we cannot legitimately refer to ‘agents’ or ‘reasons’ without knowing something about who the agent is and whether the reason resonates with their particular set of motives, projects and commitments. As Hollis says, there are ‘no universal standpoints.’\footnote{Hollis, 1995, p. 171.} This idea runs like a thread through much of Williams’s writing.\footnote{See Altham and Harrison, 1995; Williams, 1973a, 1981, 1985a, 1985b, 1996 and Hollis, 1995.} In the context of thinking about moral reasons, it may be read as a rejection of the idea that moral theories can give reasons binding on everyone, regardless of individuals’ motivational standpoints.

Williams faces a potential problem, however, with connecting reasons to individual motivational sets in this way. People have all kinds of motives. There can be no question about that. Name the action and almost certainly someone, somewhere, has been motivated to do it. If there are no reasons which are independent of particular motivational sets, how can we ever evaluate our reasons? How do we know if our reasons, to put it bluntly, are any good? And, in the case of moral reasons, how do we know if our reasons are the right ones? We need a way of justifying our motives. Williams denies that there are any independent standards of justification; there is only that which motivates. He does, nonetheless, claim to have an answer to this problem, the problem of justification. It lies in the idea of \textit{reflective deliberation}. An internal reason, he says, is one which would motivate the agent if the agent deliberated properly.
Williams’s account of deliberation is designed to claim the middle ground between two extreme positions about the nature of reasons. At one extreme is the idea that there are reasons that apply to all agents and which can be discovered by theorising. At the other extreme are reasons which agents have by virtue of what motivates them in an immediate, transparent and unreflective way. One extreme is determined by theory, independently of agents’ motives. The other extreme is determined by the here-and-now of agents’ motives, independently of theory. The first extreme is the extreme of external reasons and the second is the extreme of internal reasons. Williams’s theory of internal reason lops these extremes. On his account, they do not count as reasons; just as theory alone cannot provide reasons, neither can faulty, unreflective motivation or motivation dependent on false beliefs provide reasons. Reflective deliberation is an attempt to capture the best of what theory can provide – independence and impartiality from our immediate and often faulty motives – while still maintaining a link with the individuality of particular agents.

Unfortunately for Williams, his reflective deliberation does not quite do all the justificatory work he thinks it does.\(^6\) External reasons are necessary, I shall argue in this chapter, because they provide a standpoint outside of an individual’s motivating set from which the motivating set may be criticised or justified. But external reasons are not completely detached from internal reasons. External reasons are themselves limited by what it is plausible to have as an internal reason. This means that what can count as a justificatory external reason is limited by what it is plausible for someone to hold as an internal reason. That is, justificatory reasons must be able to motivate, and motivational reasons must be capable of being justified.

This relationship between internal and external reasons has implications for moral theories, or so I shall argue. Just as justificatory reasons must be able to motivate, so too moral theories, which ground the justificatory reasons, must be able to provide motivational reasons. Thus, the test of a moral theory is whether it could plausibly motivate. This test I have referred to as ‘the motivational test’ of moral theories. The purpose of this chapter is to set up the motivational test. Williams’s paper on internal and external reasons provides the means of doing so. The chapter has four sections. In the first section I explain Williams’s account of internal and external reasons. In the second section I show why his argument that there are no external reasons is

---

\(^6\) Williams’s paper has generated considerable discussion, and has not been without its critics. See Blackburn, 1998, pp. 264-266; Cohn, 1986; Dancy, 1993, Appendix 1; Hollis, 1987, Ch. 6 and 1995; Hooker, 1987; Korsgaard, 1986a; McDowell, 1995; Millgram, 1996; Scanlon, 1998, Appendix; Williams, 1995.
mistaken. In the third section I show what benefits are to be gained by retaining the idea of an external reason and discuss the relationship between internal and external reasons. In the fourth section I show how this relationship makes plausible the motivational test.

### 1.1 Williams’s distinction

According to Williams, an internal reason always exists relative to an agent’s *subjective motivational set*, $S$. The simplest way of understanding this relationship would be what Williams calls the sub-Humean model, according to which ‘$A$ has a reason to $\phi$ iff $A$ has some desire the satisfaction of which will be served by his $\phi$-ing.’ \(^7\) Williams’s own version of internal reasons is rather more complex than this. He revises the sub-Humean model in three important ways. First, he takes a broad view of what is contained within $S$, that is, of what motivates an individual. He includes more than simply desires:

\[
S \text{ can contain such themes as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent.} \(^8\)
\]

Second, $S$ is not static. \(^9\) It can change over time, being subject to revision through deliberation and imaginative reflection. It includes, for example, reflecting on which element within $S$ one attaches most weight to, where there is some conflict, and revising elements of $S$. The agent may come to see that he has a reason to do something which he thought he did not and arrive at this conclusion by exercising his imagination better about what it would be like if the thing came about. By solitary reflection or through the persuasions of others ‘he may come to have some more concrete sense of what would be involved, and lose his desire for it, just as, positively, the imagination can create new possibilities and new desires.’ \(^10\) Williams admits that this process of

---

\(^7\) Williams, 1981d, p. 101.

\(^8\) Williams, 1981d, p. 105.


\(^10\) Williams, 1981d, p. 105. Williams’s account of deliberation is reminiscent of Hume’s remarks on the importance of reflecting. ‘We bring our own deportment and conduct frequently in review,’ says Hume. ‘This constant habit of surveying ourselves, as it were, in reflection, keeps alive all the sentiments of right and wrong, and begets, in noble natures, a certain reverence for themselves as well as others, which is the surest guardian of every virtue.’ (Hume, 1985, p. 276 (§ 225). First published 1751.)
deliberation is vague, but sees this as a positive feature since there is an ‘essential indeterminacy’ about what counts as rational deliberation.\textsuperscript{11} It is therefore undefined as to what the agent might come to count as a reason as a result of deliberating from his existing $S$.

Third, there are two types of internal reasons. In the first, most obvious case, the reason transparently motivates. Suppose, for example, I am thirsty. I then have a reason to fill a glass with water and drink it. My thirst, that is, my desire for a drink, motivates me. In the second case, the reason fails to motivate but only because of an epistemic failure. There are three kinds of epistemic failures: I have a false belief, I am ignorant of some relevant fact, or I am ignorant of some element within my $S$. Williams’s own example is of the first kind. He considers an agent who wishes to drink a gin and tonic, and who thinks that the bottle contains gin, whereas it actually contains petrol. The question is: does he, the agent, have a reason to mix the petrol with the tonic and drink it? Of such a case, Williams says that even though the agent \textit{thinks} he has a reason to pour the contents of the gin bottle into the tonic, he does not \textit{actually} have a reason.\textsuperscript{12} That is, it is possible to have an internal reason and not know it, and to believe mistakenly that one has an internal reason. An internal reason, therefore, does not necessarily motivate, but \textit{would} motivate the agent if she had only true beliefs and was in command of all pertinent information. This second kind of internal reason gives a degree of objectivity to internal reasons.\textsuperscript{13}

What may be said of external reasons? Williams says that external reason statements ‘can be true independently of the agent’s motivations.’\textsuperscript{14} As an example, he cites a character in a Henry James novel, Owen Wingrave, whose father wishes him to join the army, in the name of family tradition and pride. Owen has no motivation to join the army and despises everything for which it stands. In short, there is nothing in Owen’s $S$ which could lead him, through deliberative reasoning, into joining the army. If his father nonetheless insists that Owen has a reason to join the army he must mean it in an external sense.\textsuperscript{15}

\textsuperscript{11} Williams, 1981d, p. 110.
\textsuperscript{12} Williams, 1981d, p. 102.
\textsuperscript{13} A similar distinction is made by John Harsanyi between manifest preferences and true preferences. A person’s true preferences are those he would have if he had all relevant information, reasoned with great care and was in ‘a state of mind most conducive to rational choice.’ (Cited in Robertson, 1986, p. 125.)
\textsuperscript{14} Williams, 1981d, p. 107.
\textsuperscript{15} Williams, 1981d, p. 106.
The difference, then, between internal and external reason statements is that the former are true by virtue of an agent’s S, whereas the latter are true independently of the agent’s S. In the case of moral reasons, an external reason exists by virtue of some moral standard, quite independently of whether a particular agent is motivated by it (or would be motivated by it if he had only true beliefs and so on). But what would it mean for someone to have a reason which has no relation to what motivates him? Williams argues that this can never be the case: there are no external reasons. His argument, the first of two arguments against external reasons, goes like this. 16

(1) If something is a reason for action then it must be able to feature in an explanation of that action.
(2) But nothing can explain an agent’s actions except that which motivates him so to act.
(3) Therefore, an external reason must be able to motivate an agent to act.
(4) But the whole point of an external reason (if there is such a thing) is that it exists independently of whatever motivates the agent.
(5) Therefore, from (3) and (4), there are no external reasons.

He considers the following defence of external reasons. In the case of the external reason, it is not the reason, R, itself which motivates the agent, but his coming to believe that R is a reason. The agent acquires the motivation because he comes to believe the reason statement and he does that because he is considering the matter aright. When I understand that something constitutes a reason I am thereby automatically motivated, whatever my original motivations may have been. 17 But how exactly do we traverse this gap between not seeing something as a reason – and not being motivated – to seeing something as a reason – and thereby being motivated? Williams says the gap would be bridged by rational deliberation. The claim about external reasons, then, must be that ‘if the agent rationally deliberated, then, whatever motivations he originally had, he would come to be motivated to φ.’ 18 To believe I have a reason is to believe that if I rationally deliberated I would be motivated to φ. This argument, says Williams, is implausible. Here is the second of Williams’s arguments against external reasons. 19

(A) Having an external reason to $\phi$ means that if an agent rationally deliberated he would be motivated to $\phi$.

(B) To deliberate, the agent must have something to deliberate from, namely his original motivational set.

(C) But if he can get to the new reason from his original motivational set, then he has, in fact, an internal reason of the second variety outlined above.

(D) Therefore, the external reason was really an internal reason all along.

What are we to make of Williams’s understanding of internal and external reasons? Let us grant Williams that, for a fact or belief to count as an internal reason for or against an action, it must engage the agent in some way; it must speak, to use John McDowell’s metaphor, to whatever it is which motivates the agent. This point explains why the giving of advice is often such a futile exercise. When giving advice to another, we advocate they $\phi$ because of reasons X, Y and Z. But unless X, Y and Z appeal to something which already motivates the person, he or she will remain unmoved. Our advice goes unheeded. Suppose we advise a teenager to attend her school studies better so that she may please her elders. If she despises her elders, we can only expect our advice to be disregarded. Pleasing her elders just does not count as a reason for her.

What are we to make of Williams’s claim that there are no external reasons? He has given us two arguments against there being any external reasons. In the following two sections I take issue with both of them.

1.2 In defence of external reasons: seeing matters aright

In this section I consider the second of Williams’s arguments against external reasons, namely that which is given in (A) to (D) above. The key premise is (A). We need to be clear about the role of rational deliberation in this premise. Williams says that an agent having an external reason must mean that the agent would be motivated by it if he were seeing matters aright. That is to say, the agent lacks the relevant motivation because he does not see things properly; he has a faulty understanding. So how could he get to the point where he did see matters aright? According to Williams, rational deliberation does the work: if

---

20 McDowell, 1995, p. 68.
the agent rationally deliberated, he would come to see that he had a reason, that is, he would be motivated.

The problem lies in Williams supposing that rational deliberation is the only way of coming to ‘see matters aright.’ If this were the case, then Williams’s point – that there are no external reasons – would be well made. For, in respect of (B) and (C) it is surely right that we cannot be led through deliberation to acquire a new motivation if the deliberative process does not appeal to some motivation which already exists. When we deliberate, we deliberate from a certain standpoint, namely the standpoint given to us by our S. Let us grant Williams that point.

But why should we assume that rational deliberation is the only means available to us of coming to see matters aright? In his paper, ‘Might there be external reasons?’ McDowell challenges Williams on this very point. Let us now consider McDowell’s argument and defend it against objections which Williams could raise.

1.2.1 The idea of conversion

How else, other than by rational deliberation, could someone make the transition from not seeing matters aright to seeing matters aright? Williams acknowledges that one could be moved by rhetoric, or other non-rational methods, but adds that these means are of the wrong kind to interest the external reasons theorist.21 McDowell disputes this point. He puts forward the idea of coming to see a matter aright through a conversion.

... there is no implication ... that there must be a deliberative or rational procedure that would lead anyone from not being so motivated to being so motivated. On the contrary, the transition to being so motivated is a transition to deliberating correctly, not one effected by deliberating correctly; effecting the transition may need some non-rational alteration like conversion.22

The issue here is the mechanism by which it is possible to move beyond one’s existing motivational set to acquire new motives. McDowell’s point is that rational deliberation is not the only valid mechanism. If this is so, there is room for explaining how one could come to believe an external reason without needing to refer to one’s existing motivational set. McDowell says that

---

22 McDowell, 1995, p. 78.
there are some individuals who have the sort of S from which it simply will not be possible to traverse the gap to ‘seeing matters aright.’ In such cases, there may be no point of leverage for reasoning which is aimed at generating the motivations which are required. For such individuals the only way for them to acquire the change of motivation is through a non-rational process. McDowell defines a conversion as just such a process, namely

an intelligible shift in motivational orientation that is exactly not effected by inducing a person to discover, by practical reasoning controlled by existing motivations, some internal reasons that he did not previously realise he had.

It suits Williams’s purposes to equate ‘coming to see matters aright’ with the process of deliberation, since deliberation requires an appeal back to the agent’s existing motivational set. Once we are back in the motivational set, we can only talk about the internal reasons which one already has. But McDowell’s point is right. Rational deliberation is not the only way to adopt new motives. We may be moved by eloquence and oration, by stories and music, and by the example of others. It may be unusual to be moved by such means without any appeal at all to our existing motives, but it is not inconceivable.

Let us make McDowell’s point plausible by considering a detailed example of a conversion. I have in mind a scene from Leo Tolstoy’s novel, Anna Karenina. After her confinement with Vronsky’s child, Anna is at death’s door. She telegraphs her husband, imploring him to return and forgive her for her affair with Vronsky. When he is summoned Karenin begrudgingly returns to Petersburg from a sense of duty. He hates his wife and wishes her dead for the humiliation and rejection she has wrought on him by her relations with Vronsky. Yet, when he sees her so ill, he is unforeseeably overcome by her fragility and suffering, and is filled with compassion. He tells her he wholeheartedly forgives her everything.

I take this incident in the novel to be a case of McDowell’s non-rational conversion. The contention here is that no deliberation could have brought Karenin to the point of forgiveness. He was just suddenly inspired by Anna’s

---

25 The magnanimity of his forgiveness, in turn, forces Vronsky, who is present, to see Karenin and indeed his own life in a new light. Vronsky himself undergoes a conversion: ‘All the habits and rules of his life that had seemed so firm, had turned out suddenly false and inapplicable.’ (Tolstoy, 1977, p. 406. First published 1875-1877.)
situation at the time of seeing her. In the scenes leading up to this deathbed scene, we gain a detailed picture of Karenin, and what motivates him, which brings home to the reader just how remarkably unexpected the forgiveness scene is. We see that he is a well-known, high-ranking public official with a strong sense of duty, order and propriety in both his public and private life. He detests displays of emotion. Further, ‘to put himself in thought and feeling in another person’s place was a spiritual exercise not natural to Alexey Alexandrovitch. He looked on this spiritual exercise as a harmful and dangerous abuse of the fancy.’\textsuperscript{26} When Anna confirms to him her love for Vronsky, Karenin desires ‘that she should get due punishment for her crime. He did not acknowledge this feeling, but at the bottom of his heart he longed for her to suffer for having destroyed his peace of mind – his honour.’\textsuperscript{27} Later we read of his fury with her for not observing proprieties as she carries out the affair.

What is more, just prior to the forgiveness scene, Karenin explicitly rules out the possibility of forgiveness when Anna’s sister-in-law and friend, Dolly, tries to coax him into a more conciliatory frame of mind. She offers two reasons for a change of heart on his part. First, forgiveness is required of him on account of his Christian faith; and second, forgiveness will make a reconciliation possible. She points out that her own marriage had suffered infidelities by her husband but all had been put right by her willingness to forgive. In fact, her forgiveness had enabled her husband to realise better the wrongness of the infidelity. Karenin’s reply? ‘Forgive I cannot, and do not wish to, and I regard it as wrong. I have done everything for this woman, and she has trodden it all in the mud to which she is akin.’\textsuperscript{28} Reflecting on this conversation later, he feels annoyance at Dolly’s suggestion of forgiveness. Yet, in the scene following this interview with Dolly, the deathbed scene, he is overcome by forgiveness.

Everything in Karenin’s S suggests that he had no (internal) reason to forgive. He believed, with some justification, that he had been humiliated and treated with contempt. He also believed Anna was wrongfully dismissing her marital duties and should be punished for it. He was motivated on account of these beliefs to divorce her. We may see Dolly’s conversation with him as an attempt to convince him, through reflective deliberation, that he did have a reason to forgive her. Yet he rejected the idea outright. Furthermore, Karenin’s unforgiving response to Anna was in character. It is what we would have expected of him, knowing him to be a man who placed great store on propriety, respect and the esteem of others. It is not as though he had temporarily

\textsuperscript{26} Tolstoy, 1977, p. 144.
\textsuperscript{27} Tolstoy, 1977, p. 278.
\textsuperscript{28} Tolstoy, 1977, pp. 386-387.
‘forgotten’ a compassionate nature. For him to forgive, to ‘see things aright’ would require a startling transition in his $S$, a transition which could only be brought about, it seems, by a McDowell-like conversion, and this is exactly what occurred.

But more needs to be said about this transition. Karenin’s conversion has three temporally related parts: the pre-conversion state, the conversion experience and the post-conversion state. The pre-conversion state is marked by a set of beliefs and a corresponding pre-conversion $S$. He was motivated by resentment towards Anna. He was unforgiving and bent on divorce. The conversion experience itself, the witnessing of Anna, is sharp, quick, and unexpected. It is marked by an intense emotional experience; he is filled with compassion and ‘the bliss of forgiveness.’ This experience brings about a change in his $S$. He forgives her. His previous malevolent motives disappear. Now he wishes to protect Anna and drop the divorce.

In the example Karenin clearly changes from being unmotivated to forgive to being motivated to do so. But what is the precise nature of this change? There are three ways we might understand the work which the conversion experience performs. We can express these three possibilities in terms of reasons.

Let us suppose Anna’s suffering is the reason Karenin has to forgive. I suggest that in the pre-conversion state this was an external reason. For him to forgive, three things have to happen. He has to:

1. have a belief that she suffers,
2. have a belief that suffering is a reason to forgive,
3. be motivated by the reason.

On the first interpretation of the conversion, if Karenin has (1) he automatically has (2) and (3). In the pre-conversion state, however, he does not have (1). That is, he does not believe that she is suffering. He might say to himself something like, ‘If I thought she was suffering, I could forgive her. As it is, she doesn’t suffer; I’m the only one who is suffering.’ The conversion gets him to (1). He acquires a new belief about Anna. He can then forgive her.

On the second interpretation, he has (1) in the pre-conversion state, but (1) and (2) are separated. Also, if he has (2), he automatically has (3). In this case we can imagine Karenin reflecting after his conversation with Dolly, ‘It is true that Anna suffers. But so what? She brought the whole thing on herself and she deserves to suffer.’ That is, he acknowledges her suffering, but it leaves him cold; he does not believe that it counts as a reason to forgive her. The
conversion experience enables him to see that her suffering is a reason to 
    forgive; (1) and (2) become connected. On the first interpretation, he learns that 
    Anna suffers. On this second interpretation, he already knows that Anna 
    suffers, but he learns that her suffering is a reason to forgive.

    There is a third interpretation of the conversion. In the pre-conversion 
    state, he has (1) and (2) but (2) and (3) are pulled apart. This amounts to 
    Karenin acknowledging he has a reason to forgive but remaining unmotivated by 
    it. With the conversion he comes to be motivated by an existing reason.

    The question now is, which interpretation best describes a conversion 
    of the type experienced by Karenin? To begin with, we can rule out the third 
    possibility, since Williams and McDowell both assume in their discussion of 
    external reasons that if the agent comes to believe he has a reason, he will be 
    motivated by it. The problem which the conversion idea addresses is not how 
    someone can come to be motivated by an external reason once they believe they 
    have it. The problem is, rather, how they can come to believe they have the 
    reason in the first place.

    We are left with two possible explanations of what happens when a 
    conversion takes place. Let us put the two options as starkly as possible. 
    Either we adopt a new belief about the world, or we come to understand that a 
    pre-existing belief has motivational pull which we had not previously realised.

    At this point I wish to leave it open as to which of the two possibilities 
    explains the conversion. I see no reason why a conversion may not operate 
    differently in different situations, now enabling a new belief to take effect, and 
    now enabling an existing belief to have cogency. For the moment, it is sufficient 
    to note that the importance of the conversion, as described here, is that it 
    shows how it is possible for a motivational set to change without such a change 
    having to appeal to what is already present within it. Whether the mechanism 
    is by instilling a new belief or by activating an existing dormant belief is a 
    secondary question.

    Before addressing the objections which might be raised against the idea 
    of ‘coming to see matters aright’ through non-rational means, let us summarise 
    what has so far been argued. In discussing the plausibility of external reasons, 
    Williams and McDowell both see the relevant question as being how a reason 
    which is not understood as such by a particular agent can come to be believed 
    by that agent and thus motivate. Williams says it can only come to be believed 
    if it is accessible, through rational deliberation, to the agent’s S. McDowell says

30 McDowell mentions these two possibilities in passing. McDowell, 1995, n. 5.
31 A further explanation is that both of these changes occur.
it can come to be believed through non-rational means. But the question of the means by which a reason comes to be believed is a separate question from its existence in the first place. Williams conflates these two questions, arguing that because the means by which it comes to be believed necessarily appeals to internal reasons, there are only internal reasons. McDowell, on the other hand, rightly argues that because the means need not appeal to existing internal reasons, the external reasons still stand.

McDowell postulates an alternative way of coming to believe a reason. A situation just assuredly strikes you from the blue, so to say, in the right way. Despite your S, you are converted to seeing you now have a reason. The example I gave of such a conversion was one from Anna Karenina. Although this is a fictional example, Karenin’s forgiveness of Anna does not strike the reader as contrived. It rings true. McDowell is right in thinking that changes can occur through non-deliberative means.

1.2.2 Objections

To make sense of the Karenin example Williams needs to provide an account of what happened to Karenin without referring to the idea of a conversion. There are two ways of providing such an account. The first is to say that Karenin had an internal reason all along, but he just did not realise it. The second is to say that in the pre-conversion state he did not have a reason at all.

(i) Having an internal reason but not being aware of it

Might it not be that Karenin had an internal reason all along, but, during what I call the pre-conversion state, he was unaware of it. There are two related possibilities as to why he may have been unaware of it. One is that he simply did not deliberate well enough. The other is that he just had a false belief and once that false belief was rectified, he was duly motivated. Let us consider the first possibility, that he could have discovered his reason through reflective deliberation.

We recall that Williams says internal reasons can be discovered in deliberative reasoning. Might it not be the case that if Karenin had just thought a bit longer and harder about the whole matter, he would have come to realise that he had a reason? That is, might not the reason have been accessible to his S all along? I have already suggested that Dolly’s attempt at persuasion was an attempt at deliberative reasoning and it failed. Is it simply that she did not try
hard enough? No. We have to realise that what can be reached through deliberative reasoning has its limits.

One limit is the person's $S$. One's capacity to reason is not entirely independent of one's $S$ itself. There are certain conditions which need to be met before deliberative reasoning can take place at all. One must be open to possibilities, have a clear head, be willing to revise judgement and have a certain humility. These conditions may not be met if one has certain strong desires, emotions or dispositions. Karenin's existing motives were such that he was not open to what Dolly had to say. His $S$ was so constructed as to preclude the possibilities which Dolly raised. This point is a facet of precisely the point that Williams wants to bring home: that argument itself cannot motivate; it must hook into existing motives. Some arguments are precluded just because the existing motives render the person 'blind' to them. Karenin's motives were a case in point.

Another limit is simply the capabilities of the person at a more general level. Williams speaks about the importance of imagination in connection with deliberation. Certainly, this was one capacity in which Karenin was limited. We know that empathising did not come naturally to him and that he believed it to be a dangerous exercise. He did not have the kind of imaginative disposition which would have enabled him to place himself in Anna's situation. Nor did he have the imagination to foresee how things might be different if he did forgive, which would have helped him to accept Dolly's second point. His austerity may have been a failing on his part, but it was true of him nonetheless. Because of his limited capabilities in this regard, his capacity to engage in critical reflection was likewise limited. This means, in turn, that his range of internal reasons was also limited.

These limits to reflective deliberation make it difficult to know whether anyone really does have an internal reason or not. How are we to know in advance whether someone is able to reflect in the way required? And, indeed, what one person could not reach by reflection, another person could, depending on what prior dispositions and capabilities the person already had. Another individual in Karenin's situation, though just as angry at Anna and repelled by her actions, may have been able, in the calmness of his study, through the smoke of his cigar, to reflect soberly on Dolly's words and see that he had reason to drop his intention to divorce and adopt a more forgiving outlook towards Anna. Thus, through reflection, elements in our $S$ may be revised, but at the same time our capacity to reflect is limited by elements in our $S$.

We conclude that Karenin was incapable of becoming motivated to forgive through a process of deliberation. The issue raised here of the limit to
reflective deliberation raises further issues about the nature of such a process. Williams is vague on this point, saying there is an ‘essential indeterminacy’ in what can be counted as a deliberative process.\(^{32}\) He also acknowledges that ‘there are no fixed boundaries on the continuum from rational thought to inspiration and conversion.’\(^ {33}\) Indeed, these elements may be almost inseparable. Karenin-like conversions — instants where an idea or a conviction suddenly dawn on us — may inter-mingle with imagination and reflection. Because of the indeterminacy of the process, it can be difficult to know whether or not someone has an internal reason.

Overall, in explaining how motives change, Williams overemphasises the role of reason, rationality and reflection. The process of change may take place through armchair deliberation but it commonly does not. Sometimes, far from change occurring because of a conscious process of deliberation, change occurs \textit{unbeknownst} to us. Until it suddenly occurs to us that we no longer want the same things we used to. Whereas at twenty, Bunbury wanted fast cars, late nights and plenty of champagne, at forty these things no longer motivate him. He finds them unsatisfying and he yearns for a more settled life. There was no intentional deliberation; at no point did he sit himself down and go through a motivational audit. Rather, the change was incremental, like the ageing process itself, a gentle maturing. Bodily changes can bring about changes in our \(S\). A prolonged illness, pregnancy, a brush with death, all these things may make us want new and different things, may make us interpret those around us differently. None of this is to deny that reflection can do the things Williams says it can. And it is not to deny that there is an important place for reflection in change. It is just to say that there are many paths to the door of ‘seeing things aright.’

One reason why non-rational ways of changing are excluded by Williams may be that they do not, in themselves, contain any normative aspect. It might be said that only deliberation can change elements of the \(S\) in a way which is purposely better. That is, only rational reflection can \textit{intentionally} challenge the merit of what motivates us. After all, Williams is not interested in change \textit{per se}; he is interested in change for the better, change to the \textit{right way of seeing things}. He is mistaken, however, in supposing that the change to the right way of seeing things can only legitimately be brought about by \textit{intending} to bring oneself to the right way of seeing things. There are any number of

\(^{32}\) Williams, 1981d, p. 110.
\(^{33}\) Williams, 1981d, p. 110.
ways we can be brought to the right way of seeing things, as I have argued in
this section.

I said earlier that there were two ways in which we can make sense of
the idea that Karenin had an internal reason all along. The first was that the
reason was indeed accessible to his $S$. The second was that Karenin simply had
a false belief, which, once corrected, enabled him to be motivated. We recall
that, according to Williams, it is possible to have an internal reason and not
realise it. One reason for this is that we may have a false belief. Was this not
the case with Karenin? If Karenin had understood just how much Anna was
suffering, he would have realised that he had a reason to forgive. According to
this interpretation, he had a false belief in thinking that she was not suffering.
Once this misconception was cleared up, he forgave. If this is correct, then he
always had an internal reason. He just did not know it because of his false
belief.

Suppose we were to concede that Karenin had a false belief. That would
not mean that he always had an internal reason. Not all false beliefs mask
internal reasons. We have to consider the reason why the person has a false
belief. In Karenin’s case, an accurate belief about Anna’s suffering was only
accessible to him via the kind of conversion experience he had. He could not, as
we have seen, have acquired the belief through deliberative reflection, and for
the very reason that his $S$ got in the way. We have already made the point that
his capacity to reason was limited, blinded if you like, by his temperament. He
was too rigid, too angry and jealous to see that Anna was suffering. He had to
step outside his $S$ for an instant to acquire the belief. For it to be the case that
we have an internal reason of which we are unaware on account of a false belief,
the false belief must be of the right origin. If we have the false belief because of
the $S$ itself, we can scarcely be said to have an internal reason. If the reason has
to do with a simple lack of information or deceptiveness on the part of others,
which, if we knew about would not be in tension with our $S$, then it is fair to
say we have an internal reason. We have to put this restriction on internal
reasons or the concept becomes too detached from the idea of an $S$, an idea
which is exactly what Williams wishes to preserve.

This rather technical point is critical to an understanding of the scope of
internal reasons. The important feature of internal reasons is that they are
limited by a person’s motivational set: reasons are contextualised to motives.
We can agree with Williams on that point. However, some things will never be
seen as reasons because a person’s motivational set blinds them. I hope to have
shown this through the example of Karenin. How are we to understand this
‘blindness?’ If we take it simply to mean ‘false beliefs,’ it might be thought
that the person could still have an internal reason which was blocked by the false belief. In this case, the person has an internal reason but does not know it because of his or her false belief. The problem with this way of looking at things is that 'blindness' means a great deal more than simply having false beliefs. It means being unable to acquire the relevant true belief because of the restricting nature of the S. As we saw, Karenin was too unsympathetic ever to be able to appreciate Anna's suffering and remorse. He could not acquire the relevant true belief with his existing S. Because of this he did not have an internal reason. To say otherwise would be to deny altogether the relevance of existing motivational sets and that is to deny the very point that the idea of internal reasons sets out to capture.

(ii) Having an external reason or having no reason?

At the outset of this sub-section I said there were two replies that Williams could make to the Karenin example. The first was that Karenin had an internal reason all along. That possibility has now been shown to be implausible. We are now in a position to consider the second reply which is open to Williams. Williams could say that, prior to the conversion experience, Karenin did not have a reason to forgive Anna at all. He did not have an internal reason and he did not have an external reason. With the conversion, his motives changed and he came to have a new (internal) reason.

We can agree with Williams that, prior to the conversion experience, Karenin did not have an internal reason. I contend, however, that at that time Karenin had an external reason. That is, he had a reason even though it did not motivate him. It makes sense for Dolly to have said to Karenin, 'There are good reasons why you should forgive Anna,' even though Karenin could not see the reasons. Williams would simply deny this point. He would need to say that it was not possible for Karenin to have a reason since Karenin's S was not constituted in the right way. In saying that Karenin had no reason at all, Williams is simply stipulating that there are no external reasons.

There is now something of a stand-off between these two points of view. However, by allowing for external reasons, we can explain the common parlance. For, as I remarked above, it seems fitting that Dolly should refer to reasons in her dialogue with Karenin. It also seems fitting that we say of Karenin that he was unable to understand that he had reasons to forgive. This is a better way of explaining the situation than saying that he had no reasons to forgive. If that were the case, what did Dolly think she was doing when she tried to persuade him to forgive Anna? Indeed, what is persuasion other than an
attempt to enable another person to see that they have reason for a certain action? In the next section I show what can be gained by retaining the idea of an external reason.

1.3 In defence of external reasons: the justificatory standpoint

In the first section we identified two arguments which Williams presents against external reasons. The previous section addressed the second of these arguments, so let us now return to the first of them. Here it is again.

(1) If something is a reason for action then it must be able to feature in an explanation of that action.

(2) But nothing can explain an agent’s actions except that which motivates him so to act.

(3) Therefore, an external reason must be able to motivate an agent to act.

(4) But the whole point of an external reason (if there is such a thing) is that it exists independently of whatever motivates the agent.

(5) Therefore, from (3) and (4), there are no external reasons.

Let us consider carefully premise (1) since, if we accept this premise, then the rest will plausibly follow.34 Reasons, says Williams, must be able to explain actions. This is obviously true for internal reasons, which, by definition, are connected to motives, but need it be true for all external reasons? No, it need not. We can agree with Williams that an external reason would motivate the agent if the agent saw matters aright. In the previous section we considered what would be involved in getting the agent to that point. We established that a conversion along the lines experienced by Karenin was a plausible possibility. But there is another possibility altogether: our agent may

34 Cohon would disagree with me about everything following from premise (1). She takes issue with premise (2). Many actions she says, can be readily explained even though there is no element of the agent’s S that the action furthers. For example, ‘on a particular evening a man makes a sack lunch just as he does every evening; he knows that tomorrow he plans to lunch in a restaurant with a colleague, but as he performs his routine task he does not happen to make the connection.’ (Cohon, 1986, pp. 548-549.) However, this example doesn’t negate Williams because it could be understood in terms of the man having a mistaken reason. That is, he thinks he has a reason but he has not. He has forgotten a relevant fact, namely, that he will not be needing a sack lunch. This example is not different in structure from the case of the gin and tonic, where the agent thinks he has a reason to pour the substance into the tonic, but he is mistaken.
never be motivated to seeing things aright. In this case, he never makes the reasons his own. To be sure, this would indicate a flaw in his reasoning. But that is not a problem. As Christine Korsgaard reminds us, ‘rationality is a condition that human beings are capable of, but it is not a condition that we are always in.’

But is there any point in retaining the notion of an external reason if we acknowledge that it is not necessarily connected with motives? I think so. External reasons have a function independently of whether they motivate, since external reasons provide a justificatory standpoint. They enable us to appraise another’s actions and (explanatory) reasons for actions from a standpoint outside of that other’s motivational set. This section defends the justificatory role of external reasons.

1.3.1 The role of external reasons

Williams distinguishes between two kinds of conceptual work which an internal reason does. A reason, he says, ‘is concerned with the agent’s rationality.’ But a reason also provides an explanation of action. That is, a reason tells us whether an action is rational and it explains why an action was performed. These two functions may sometimes conflict. We recall Williams’s example about the person who goes to mix a gin and tonic. The person thinks that the liquid is gin but it is really petrol. According to Williams, this person does not have a reason to pour the liquid. He only thinks he has a reason. However, once the liquid is poured, we may always ask the question, ‘why did he do it?’ We will be able to give an explanatory reason after the event. Therefore, before the action, we say, according to Williams, that he has no reason to drink the glass’s contents, but after the action, we say that, of course, he had a reason to drink the petrol. This seeming contradiction can be explained by the twofold kind of work that reasons do, namely, justify according to criteria of rationality or morality, and explain. The agent has a reason for drinking the liquid in the explanatory sense of a reason, but not in the rational or justificatory sense of a reason.

Yet, Williams does not allow that external reasons may also be justificatory. We recall that the first premise of the argument that there are no external reasons is that ‘if something can be a reason for action, then it could be someone’s reason for acting on a particular occasion, and it would then figure in

35 Korsgaard, 1986a, p. 18.
36 Williams, 1981d, p. 103.
an explanation of that action.\textsuperscript{37} To put it another way, an external reason for an action \textit{must be able to figure in an explanation of that action}. Because explanations are always retrospective, that is, refer to what has already occurred, the problem then becomes one of providing a link between an external reason and motivation. But the first premise is problematic in the following sense. Providing an explanation \textit{is not the only kind of work which a reason can perform}. We have seen that a reason also has a justificatory function. Even if an external reason cannot do explanatory work, that does not rule it out as being capable of justificatory work.

To summarise, by insisting that external reasons need to be able to explain action, Williams presents an account of external reasons which is doomed to fail. Of course external reasons cannot be explanatory. To explain action they must be able to motivate. Yet the whole point of external reasons is that they have no connection to motivational sets. The point here is that external reasons have a justificatory role, not an explanatory role. They provide the means by which we can properly say that, although someone had no (internal) reason to \(\phi\), he \textit{ought} to have had. The external reason permits us to escape the perspective of an agent; it enables us to make an evaluation of the agent’s actions from an external point of view. Williams has not argued against external reasons being justificatory.

Reasons perform (at least) four kinds of work. They may best be understood in terms of the following two questions which we may ask concerning actions.

(i) What is going to happen?
(ii) What did happen?

That is, we can ask for a prospective or retrospective account of actions. The prospective question, (i), may be answered in two ways: firstly, what is going to happen by the lights of the agent’s motivational set and secondly, what ought to happen according to some \textit{independent standard} of rationality or morality.

Similarly, the retrospective question, (ii), may be answered in two ways. Firstly, what actually happened, according to the agent’s motivational set, and secondly, what ought to have happened according to some independent standard of rationality or morality. Accordingly, (i) and (ii) may be expanded into four questions.

\textsuperscript{37} Williams, 1981d, p. 106.
(i) What is going to happen according to \( S \) (or what would happen if the agent had true beliefs, full information and so on)\(^{38}\)
(ii) What ought to happen, according to the independent standard?
(iii) What happened, according to \( S \)?
(iv) What ought to have happened, according to the independent standard?

We need to clarify these four questions even further. Question (i) requires an answer in terms of the agent’s actual motivational set. Question (ii) is a justificatory question. The answer will tell us what motivation the agent \textit{ought} to have if he is to satisfy the independent standard. Question (iii), when answered, gives us an explanation of the action in terms of what actually did motivate the agent. And question (iv), similar to (ii), will tell us, when answered, what motivation the agent should have had to satisfy the independent standard. It can tell us whether the agent was justified in his motivation.

We are now well placed to see what work is done by internal and external reasons. An internal reason performs the functions of (i) and (iii). It tells us what reasons the agent has or had, relative to his \( S \). I take this to be uncontroversial. An external reason performs the functions of (ii) and (iv). It tells us what ought to motivate or ought to have motivated the agent, according to some point of reference outside of the agent’s \( S \). Williams’s mistake, in arguing against external reasons, is to think that (i) and (iii) are the only functions a reason performs.

We do, in fact, refer to ‘reasons’ in all four senses, which accounts for the complexity of the term. There is a multifarious relationship between reason and motivation. Contrary to Rachel Cohan, who refers to internal and external reasons as being ‘mutually incompatible,’ nothing precludes a reason performing more than one function.\(^{39}\) A reason may be both (i) and (ii). That would be a desirable state of affairs since it would mean that our agent is motivated by that which \textit{ought} to motivate him. His motivation is \textit{justified}, not just by the lights of his own motivational set, but on independent grounds. If the reason is (i) but not (ii) he has a \textit{bad} or \textit{wrong} reason. He is not motivated by what \textit{ought} to motivate him.

\(^{38}\) We need to add the caveat, ‘or would happen ... ’ to cover cases like the gin and tonic example, where the agent is not properly informed. The point of contrast between (i) and (ii) is that (i) relates to actual motivational sets, whereas (ii) does not.

\(^{39}\) Cohan, 1986, p. 545.
We note that on this account we are able to speak of good and bad reasons because we have some external point of reference. On Williams’s account, we are only able to say whether a person does or does not have a reason, or whether he is mistaken in thinking he has a reason. In Williams’s example of the gin and tonic, the person who drinks the substance, thinking it to be gin, according to Williams, is mistaken in thinking he has a reason to drink it. A false belief points to a mistaken reason. If we allow for external reasons as I have described, a false belief points to a bad reason. The person certainly has a reason for drinking the liquid; it just is not a good reason because one crucial belief of his is false.

1.3.2 Objection

We have not quite sufficiently argued the case for external reasons. There is still a reply which might be made on behalf of Williams. He might wish to say that the functions which are putatively carried out by external reasons can all be fulfilled by internal reasons. That is, there is no need to postulate external reasons just so that a justificatory role may be performed. An internal reason, it might be said, also performs (ii) and (iv). We recall that an internal reason may not actually motivate an agent. If it can be shown that the person ought to be motivated by what becomes apparent through rational deliberation, then that is sufficient for there to be an internal reason as well. Therefore, goes the reply, the internal reason is also concerned with what ought to motivate.

But if all this is so, Williams’s point about reasons having to be situated within the context of an agent’s motivational set loses bite. The internal reason is looking pretty external, on this line of reasoning, since it is cast adrift from considerations of what actually does motivate the agent, or, more to the point, what actually would motivate the agent if he did deliberate. The test, in the end, of whether something mooted as a reason would motivate a person if he deliberated is whether it does, in practice, motivate when he deliberates in that way. This is something we cannot always tell in advance. Williams’s discussion of those internal reasons which fail to motivate pivots on a very simplistic example about gin and tonic. This example rides on an important assumption: that if the person corrected his false belief he would change his motivation. But let us suppose he is motivated to drink petrol, because he has a death wish, or has some other quirky element in his S. In such a case, it turns out that he still has a reason to drink the substance. There is nothing to prevent our man having desires which give rise to improvident action, provided the desires are, as Hollis
says, ‘assembled with consistency and judgement.’\textsuperscript{40} If he understands the effects of drinking petrol and tonic but still wants to do it, he will be acting rationally when he drinks the concoction.

The point here is that an internal reason is always contingent on what the individual’s (possibly flawed and unfathomable) $S$ happens to be. As Hollis points out, there can be no further complaint about the desires of an agent who has achieved ‘a self-conscious reflective consistency. There is no more external sense in which weights can have been wrongly assigned to existing motivations.’\textsuperscript{41} Internal reasons cannot be contingent on some ideal set of motives, the perfectly rational $S$. The notion of an $S$ cannot be manipulated in this way, since to do so is to lose what is so useful about the idea of linking reasons to particular motivational sets. What is useful about Williams’s account of internal reasons is precisely that it does draw our attention to the fact that (at least in some cases) reasons have to be contextualised to particular individuals with particular motives. If it turns out that what this means is that reasons have to be contextualised to a perfectly rational ideal $S$, the usefulness of referring to individuated motivational sets at all, is lost.

Williams finds himself on the rather sticky horns of a dilemma. He can cast the net of reflective deliberation so widely that internal reasons fulfil a justificatory function, in which case internal reasons lose their relevance to motivational sets, or he can keep the relevance to motivational sets and lose the justificatory function of internal reasons. Much simpler just to admit that there are external reasons which fulfil the justificatory function.

\subsection*{1.4 The motivational test}

We need to be clear about what has been argued so far and what has not. We have established that Williams’s arguments against external reasons do not work. In the process two points have been made. One is that external reasons perform a justificatory role which cannot be adequately carried out by internal reasons. The other point is that, while we can agree with Williams that an external reason would motivate an agent if the agent were ‘seeing matters aright,’ it may well be that, because of the contingencies of the initial motivational set, the agent never comes to recognise an external reason.

\textsuperscript{40} Hollis, 1987, p. 85.
\textsuperscript{41} Hollis, 1987, p. 79.
Alternatively, the agent may only come to do so by a non-rational process such as a conversion.

We now require a positive account of the nature of external reasons. We have observed that external reasons are made possible by virtue of there being an ‘independent standard’ of justification which may or may not be acknowledged by an agent. But the nature of this independent standard remains vague. Two possibilities present themselves straight away, along Kantian and Humean lines respectively. The Kantian view of the independent standard may be described as the view that rationality or morality has an authority of its own, quite independently of our contingent desires and motivations. The objectivity of the Humean’s justificatory standpoint, on the other hand, is less steadfast. On the Humean view, one is able to criticise and evaluate another’s reasons, but the criticism is always grounded in one’s own motivational set. That is to say, in evaluating others’ reasons, and declaring what everyone has a reason to do, one cannot step outside one’s own set of motivating values and adopt a universal standpoint. The Kantian externalist claims that when Dolly tells Karenin he has a reason to forgive, she is right (if she is right) because there is a reason for him to forgive. The reason is objectively grounded, quite independent of both Karenin’s and Dolly’s motivational sets. The Humean externalist claims that when Dolly tells Karenin he has a reason, Dolly is only speaking from – and can only speak from – within her own motivational set.

It is quite apparent from this sketch of the two positions that the Humean external reasons are not so very removed from Williams’s internal reasons. Like Williams, the Humean externalist denies that there are any rationally compelling reasons outside particular motivational sets. However, the Humean goes further than Williams can go in stating what counts as a reason. We saw that Williams must say that Karenin, for example, had no reason to forgive Anna. In the end, Williams’s account of Karenin’s reasons has to be confined to what Karenin could arrive at through reflective deliberation. We saw that the problem with this was that Karenin’s deliberative processes were restricted on account of his flawed set of motives. Where a person’s deliberative reasoning is flawed because of the limitations of his motivational set, Williams has to admit that the person has no reason. The Humean who grants the existence of external reasons, by contrast, may say that Karenin did have a reason even though he failed to see it. To be sure, this judgement on the part of the Humean is grounded in the Humean’s own motivational set, but this

---

is acceptable since the Humean’s deliberative processes are not flawed in the same way as Karenin’s processes.

Let us now scrutinise the Humean account of the agent who deliberates perfectly rationally. We may do so by considering an example which Williams gives in his refutation of utilitarianism.\(^{44}\) George needs to decide whether or not to accept a job researching chemical and biological warfare, to which he is opposed. For the sake of argument, suppose George accepts the position. He has misgivings about the job, but reasons that the positive consequences of taking it outweigh the negative consequences. Let us also suppose that he deliberates \textit{conscientiously}, by which is meant he deliberates with a view to finding out what the right thing to do is. He wants to make the right decision, in a moral sense. In such a case, he would be doing something more than simply reflecting on which of his motives is strongest. He is trying to match his motives with pre-existing standards of conduct. When he finally concludes that he has greater reason to take the job than to leave it, he believes that there is greater reason to take the job. His choice does not \textit{create} the reason. Moreover, his belief in there being some independent measure of his motives cannot be explained away entirely as some idiosyncratic psychological need. When George reflects, he \textit{searches} for a fact of the matter about what counts as reasons; he does not simply manipulate existing motives. As Iris Murdoch says, ‘he thinks that some things really are better than others and that he is capable of getting it wrong.’\(^{45}\) The Humean externalist is not quite externalist enough.

The Humean externalist differs from the Kantian externalist on ontological grounds. The Kantian externalist has an objective ontology; the standards – that is, justificatory reasons – exist quite regardless of anyone’s motives. The Humean externalist, on the other hand, does not believe there are reasons independent of motivational sets. But the Humean externalist differs from Williams in respect of the \textit{judgements} which can be made about the reasons others have. The Humean externalist can say that others have reasons even when those others do not themselves believe they have reasons. We have seen that Williams must say that an agent A has no reason to \(\phi\) if A’s motivational set is not of the right sort. But it is still open to the Humean externalist to pass a judgement that A does, indeed, have a reason to \(\phi\) even if A does not know it.

\(^{44}\) Williams, 1973a, esp. pp. 97-98.
\(^{45}\) Murdoch, 1970c, p. 79.
We are therefore able to make a distinction between the existence of reasons and judgements about the existence of reasons. We are also able to make a distinction between reasons that are relative to a particular motivational set and reasons which are not relative. These two distinctions give us four possibilities and enable us to plot the differences between various positions.46

<table>
<thead>
<tr>
<th></th>
<th>Subjective (relative to S)</th>
<th>Objective (non-relative to S)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ontological</strong></td>
<td>(1) reasons exist relative to particular S</td>
<td>(2) reasons exist independently of particular S</td>
</tr>
<tr>
<td>(existence of reasons)</td>
<td>X’s reasons exist relative to X’s motivational set, $S$</td>
<td>$X$’s reasons exist non-relative to $X$’s $S$</td>
</tr>
<tr>
<td><strong>Epistemological</strong></td>
<td>(3) acceptance or rejection of reasons must be judged relative to particular S</td>
<td>(4) acceptance or rejection of reasons can be rationally judged independently of particular S</td>
</tr>
<tr>
<td>(judgements about reasons)</td>
<td>$X$’s judgement and outsider’s judgement about $X$’s reasons must be made relative to $X$’s $S$</td>
<td>outsider’s judgement about $X$’s reasons are non-relative to $X$’s $S$</td>
</tr>
</tbody>
</table>

Williams subscribes to (1) and (3). This makes him a strong internalist about reasons, since he is subjective in both his ontology and epistemology.

The position we have identified as Kantian externalist has (2) and (4). This is a strong externalist position since both the ontology and epistemology do not need to refer to any particular set of motives.

---

46 This table is a modified version of that given by John Kekes in his discussion of standards for good lives. (Kekes, 1989, p. 209.)
The position we have identified as Humean externalist has (1) and (4). This position, having both subjective and objective elements, may be described as either weak externalist or weak internalist, depending on whether we give priority to the ontology or epistemology.

We now come to my own position which postulates an interconnection between (2) and (1). It is a two-fold idea:

(a) (2) provides an independent justificatory standpoint for (1), and
(b) (1) puts a limit on what (2) could plausibly be.

These two claims show how internal and external reasons are inter-dependent. Claim (a) means that internal reasons, found in category (1), are justified or otherwise according to the external reasons found in category (2). At the same time, according to claim (b), the reasons found in category (2), although independent of any particular S, cannot be independent of every S. In other words, what counts as an external reason cannot be considered independently of what could be a plausible internal reason. It would also follow as part of claim (b) that judgements about reasons have their ontological grounding in (2).

It should be pointed out that the differences between Williams, the weak externalist and my own position are not as pronounced as this table suggests. All these positions attempt to capture the middle ground between an extreme internalism and an extreme externalism. The internalist must overcome the problem that people can have flawed desires. There must be some way of evaluating and justifying the desires which people actually have. As John Robertson says, ‘there is good reason for internalists of every stripe to covet the resources available, arguably, only on the externalist side of the dichotomy.’\(^47\) Williams thinks that reflective deliberation will be sufficient for this justificatory requirement. That is to say, he thinks it will provide a justificatory standpoint without needing to invoke externalism. But we have seen that his account of rational deliberation is not sufficiently independent of individual motivational sets.

The advantage of the position I am advocating is that it takes the best of both the internalist and externalist positions. It gives a place to people’s actual motives, while still allowing for the existence of independent standards. It is what Robertson calls the ‘hybrid view.’\(^48\) Rather than opting for either internalism or externalism we are able to acknowledge both kinds of reasons by

\(^{47}\) Robertson, 1986, p. 128.
\(^{48}\) Robertson, 1986, p. 126.
postulating a necessary connection between the two. Robertson’s suggestion is that this hybrid view can best be understood by invoking different locutions for each kind of reason. One gives an externalist account of the reasons there are, but an internalist account of the reasons any given agent has.

... perhaps, the reasons there are for A to φ are just the reasons the best moral theory supplies. The reasons A has to φ are determined by psychological facts about A, his beliefs and desires. We would not have, then, incompatible theories about the nature of the relation between agents and reasons but, for all that has been shown, compatible theories about two different sorts of reasons.49

The claim, however, is not simply that there are two different sorts of reasons – internal and external – but that there is an interdependency between these two kinds of reasons. We shall see that this claim has important implications for moral theory. Let us now confine our discussion to moral reasons.

There are two parts to the relationship between internal and external moral reasons. The first is that external reasons provide a limit to internal reasons. By this I mean that what does morally motivate us is evaluated by considerations of what ought to motivate us, according to some independent standard. To put it plainly, external reasons justify internal reasons. What is it to justify? It is to show the rightness or worthiness of the internal reason. External reasons enable us to say that ‘some lives are better than others, whatever people want. We do think a contented slave has a reason to seek his freedom, whatever he apparently wants.’50 If I believe a certain course of action is the right one, morally speaking, I justify this belief, in part, by reference to what I consider to be a moral standard, which is valid independently of my desires.

The second claim is the more controversial. It is this. For a consideration to count as an external reason it must be plausible that it could be someone’s internal reason. This means that a justifying reason must plausibly be able to be a motivating one. We will presently consider what ‘plausibly’ means in this context.51 We recall that Williams said that for something to count as a reason it must be able to figure in an explanation.52 He used this point to

49 Robertson, 1986, p. 126. Robertson, who wishes to argue the externalist position, makes this suggestion in passing but does not argue against it.
50 Robertson, 1986, p. 133.
51 See §1.4.2.
52 Williams, 1981d, p. 106.
rule out the possibility of external reasons. The reply, made in the previous section, was that not all reasons provided explanations, that is, were motivating. Reasons could also be justificatory and this was the role of external reasons. And this is so: external reasons do justify. However, what may count as a justification must be something which someone, somewhere could plausibly use as an explanation for her action.

The independent standard of justification cannot be set without regard to what it is plausible for someone to have as a moral motivation. The standard is independent of any particular agent but cannot be independent of all agents. Internal reasons, thus, are assessed by external reasons and external reasons are limited by internal reasons.

The purpose of this section is to make both claims about the interdependence of internal and external reasons plausible and show how the second claim – that external reasons are limited by what it is plausible for someone to hold as an internal reason – leads to the motivational test.

1.4.1 The relationship between internal and external reasons

There are two claims to be defended. The first is that external reasons provide a limit to internal reasons, and the second is that internal reasons provide a limit to external reasons.

(i) External reasons as a limit to internal reasons

External reasons provide justification for internal moral reasons. They enable us to challenge, criticise and justify those internal reasons. External reasons are themselves justified by moral theories. It might be objected at this point that theories do not provide the standards which are independent of any particular individual motivational set, but cultural norms. That is to say, it might be said that the standards by which we judge, criticise and justify ourselves and others are set by our parents’, family’s and wider community’s norms and customs, not by the contents of Kant’s *Groundwork for the metaphysics of morals*, a work of which few outside the philosophical community have even heard, let alone read.

All this seems hard to quarrel with. Yet, a further question still remains: are these norms justified? One may have been brought up to tell the truth, but it is still open to one to question whether such a norm is a good one and if so, why. Theoretical argument lays claim to justifying or undermining such norms. Therefore, while it may be true that individuals do not need to have read the
Groundwork to know that there is an obligation to respect others, if they were to keep challenging why there are such obligations, what the source of them is, and why our community is justified in prizing an attitude of respect for others, they would in fact find themselves issuing arguments similar to those put forward in the Groundwork or Hume's Enquiry or other philosophical texts. Such works, after all, were conceived by ones who sought to explain and justify moral perspectives.

A related objection is that one does not need much theoretical knowledge to be a good person. As Kant says, there is 'no need of science and philosophy to know what one has to do in order to be honest and good, and even wise and virtuous.'\textsuperscript{53} One simply needs to follow one’s proclivities, provided those proclivities are good. Again, this is hard to deny. But if one were to ask what makes the difference between good proclivities or tendencies and not-so-good tendencies and downright destructive tendencies, one would again find oneself theorising.

Before too long, a plea for the importance of external reasons turns into a plea for the importance of theorising itself. To be rejected here is an anti-theoretical strand of thought in contemporary moral philosophy.\textsuperscript{54} The need to theorise about moral matters arises from the same need to theorise about scientific matters or religious matters. As David Stove observes, sooner or later, left to themselves, people will philosophise, ‘whatever special field of intellectual work they are engaged in, or even if they are engaged in none.’\textsuperscript{55} Humans have the capacity – and a need to use that capacity – to try to understand themselves and their place in the world at large and indeed the universe. All human societies have a cosmology, a theory about how they fit in to the wider cosmos.

I shall now consider briefly one such anti-theory understanding of ethics. This approach, as I see it, can be interpreted as being founded upon internal reasons only, unlimited by independently grounded external reasons. This discussion is intended to clarify the importance of the independent standpoint provided by external reasons.

In a paper on the moral relations between people and animals, Cora Diamond criticises Peter Singer and Tom Regan, who say that certain practices


\textsuperscript{54} See Diamond, 1991; Gaita, 1991; Rhees, 1965 and Winch, 1972 for examples of this approach. For a discussion and rejection of the anti-theory approach see Pettit, 1997, pp. 115-117.

\textsuperscript{55} Stove, 1995, p. 64.
Internal and external reasons • 43

towards animals are ‘speciesist.’\textsuperscript{56} Singer and Regan urge consistency of treatment towards humans and animals, according to the possession of certain relevant properties. Their method, says Diamond, is to take certain morally significant properties which are possessed by humans and then try to show that because those properties extend to non-humans, we should accord those non-humans a similar moral status to humans. One such property is the capacity to experience pain. Another property is rationality. She summarises their argument as follows.

If we say “These \textit{animals} are not rational, so we have a right to kill them for food,” but we do not say the same of \textit{people} whose rationality cannot develop or whose capacities have been destroyed, we are plainly not treating cases alike.\textsuperscript{57}

Not only is such moralistic ‘natter’ about speciesism obtuse, says Diamond, it is shallow.\textsuperscript{58} All it amounts to, she says, is that ‘knee-jerk liberals on racism and sexism ought to go knee-jerk about cows and guinea-pigs.’\textsuperscript{59}

Diamond has a different approach to animal ethics. Just because animals and humans share certain properties, she says, it does not follow that they share the same moral status. We need to look to the different practices we have in respect of the two groups. One difference is in how we honour the dead. In the case of humans, we have elaborate rituals and ceremonies. Such ceremonies, though they do occur (usually in affluent societies), would be regarded by most people as inappropriate, ridiculous even, when they take place in respect of animals. Another difference is whether we eat the creature in question. Eating a human, cannibalism, is quite a different practice from eating an animal. What underlies our repugnance to eating human flesh is the view that ‘\textit{a person is not something to eat.’}\textsuperscript{60} Similarly, we also respond to the human face, with its enormous capacity for variation of expression quite differently from the face of an animal. The concept of an animal is to be understood, she says, in terms of the attitudes, emotional response and practices in which we engage, rather than in a property which can be scientifically delineated and tested.

Diamond is right to notice that practices differ. But she does not tell us how this account provides a basis on which we can be \textit{critical} of our practices.

\textsuperscript{56} Diamond, 1991b. For a similar view see Gaita, 1991, pp. 117-119.
\textsuperscript{57} Diamond, 1991b, p. 320.
\textsuperscript{58} Diamond, 1991b, pp. 322, 323.
\textsuperscript{59} Diamond, 1991b, p. 325.
\textsuperscript{60} Diamond, 1991b, p. 322.
By her logic a good argument against giving political rights to women in the nineteenth century would have appealed to the view that ‘a woman is not someone who votes.’ Indeed, women did not vote and that was precisely the problem. Diamond’s account is inherently conservative since it does not allow adequate room for change, for improvement in particular. Singer’s point is precisely that past and present practices in respect of animals have been morally unjustified. This point is not as easily dismissed as Diamond would have it. To be critical of our practices, we need an external point of reference. (This is the role Williams wants to give to critical reflection.) If we are only able to consider what reasons we have relative to our existing set of motives, it is difficult to obtain a critical purchase on our practices. We need access to a theory which has claims to being true independently of what our existing motives and practices are. Hollis points out that ‘if we take seriously the demand that an agent always have an existing motivation to deliberate from, then there has to be a bias towards the present state of the self.’61 The more internalist we are about reasons, and the less willing we are to countenance an external justificatory standpoint, the more conservative we will be about the reasons we have to act. Diamond’s insistence upon the validity of existing practices (or motives) in respect of animals and humans gives her approach a strongly conservative and authoritarian bias.

For example, it is true that we react differently to the faces of animals from those of humans, but this may be partly due to an unfamiliarity with animals. Animal faces may all seem to be identical in just the same way as people from one race sometimes perceive those of another race as looking the same. This apparent sameness could signify a lack of attentiveness, interest or familiarity, rather than a profound difference in kind between the observer and the observed. Researchers working closely with animals in the wild report an awareness of individual differences between animals that would not be apparent to those lacking experience of being close to animals.62

The point here is that even if there are differences between how we ‘see’ people and how we ‘see’ animals, these differences may indicate a failing on our part, and be something it is possible and desirable to change. We need to be able to ask whether our practices regarding animals are based on good reason or on prejudice. Singer’s pointing to various properties held by humans and animals is one way of doing that. Similarly, a vegetarian could say that what is wrong with our attitude to animals is precisely that people do have the view

61 Hollis, 1987, p. 87.
62 Conversation with Michelle Hall, zoological researcher, Research School of Biological Sciences, Australian National University, May, 2000.
that ‘an animal is something to eat.’ Just because people do hold this view does not make it the right one, it could be argued, in light of the suffering involved for animals in their husbandry and slaughter. Diamond cannot have spoken the last word on what can be said about animal ethics just by pointing to certain existing practices.

But there are constraints on what can count as an external reason. It must be intelligible for the reason to be adopted. An attitude or practice prescribed by a theory cannot be so radical as to be completely outside the ambit of what could plausibly be adopted or enacted. To be critical, then, of a culture’s practices, one must draw on elements of a tradition which is already present in the culture, albeit in a minority. Much more could be said on this topic, but let us now turn to the second claim which is our main focus.

(ii) Internal reasons as a limit to external reasons

We have seen that internal reasons are justified or otherwise by external reasons. However, external reasons themselves have to come from somewhere and have themselves to be justified. What makes a plausible external reason? One necessary condition is that it could make a plausible internal reason. What can count as an external reason is limited by what it is plausible for someone, somewhere to hold as an internal reason. Someone must be able to make it her own. Thus, while external reasons are independent of a particular individual’s motivational set they cannot be independent of everyone’s motivational set.

Consider again Williams’s story about Jim and Pedro. If Jim is justified in killing the one Indian because it brings about the best consequences possible in the circumstances, then it must be plausible that Jim could be motivated to kill the one because it would bring about the best consequences possible in the circumstances. Or to take another example, if Earnest would be justified in betraying confidences and showing disloyalty to colleagues by whistle-blowing on corrupt practices in his organisation, for the reason that such whistle-blowing will bring an end to the corruption for the greater welfare of the community, then it must be plausible that he, or someone else, could be motivated to whistle-blow for that reason.\(^63\)

To put it more abstractly, internal reasons being a limit to external reasons amounts to the following principle of motivational capacity, (M).

\(^63\) This is not the same as saying that Earnest must be motivated by this reason. He may well be motivated by revenge or a personal dislike of the characters involved. He may have any number of motivations. However, it must be plausible that he could be motivated by the community’s greater welfare.
If it is right for any agent, A, to φ for reason R, then it must be plausible for some A to be motivated by R to φ.

This strikes me as a common sense requirement. It will not appear so to all, however. The objective consequentialist, for example, will either deny it is common sense or agree with Jack Smart that common sense is no great ally, being ‘in part made up of superstitious elements, of morally bad elements, and of logically confused elements.’ The objective consequentialist denies that consequentialism — as a moral theory — need have anything to say about motivation. This theory holds that its purpose is to justify certain actions, not motivate them. As Peter Railton says, ‘Objective consequentialism has the virtue of not blurring the distinction between the truth-conditions of an ethical theory and its acceptance-conditions in particular contexts.’ But whether a radical separation of these two sets of conditions is a virtue, rather than a fatal flaw, is precisely the question at issue.

The objective consequentialist — or indeed anyone who believes moral justification is a completely separate question from moral motivation — may say that we are quite familiar with justificatory reasons having nothing to do with motivation in all kinds of commonplace activity. Why, then, should it not be the case in ethics? Here might be one such example. When riding a bicycle it is not necessary to bear in mind or even know about the (Newtonian) theoretical interplay of gravity, inertia and friction. Almost everyone who rides a bicycle is ignorant of the physics behind the activity. All I need to know is that I should turn the handlebars in a certain way when going around a corner. It might be said that the various Newtonian laws of physics justify the action. But it is not plausible that this justification must be able to motivate.

Put like this, the bicycle riding example sounds like a case of the difference, made explicit by Gilbert Ryle, between knowing how and knowing what, or the difference between practical knowledge and propositional knowledge. Is this distinction the same as the distinction between justification and motivation? Not quite.

---

64 Smart, 1956, p. 173.
65 Railton, 1984, p. 155. There are actually two kinds of objective consequentialism, as we shall see in § 2.4.1. The type identified here is ‘extreme’ objective consequentialism because it denies any connection at all between moral theory and motivation.
66 This argument, together with an analogous example, is given by Eugene Bales, 1971, pp. 264-265. See also Scheffler, 1994, p. 52.
67 Ryle, 1949, II § 3.
There are important differences between scientific theories and ethical theories. It may be said fairly uncontentiously that ethics is concerned with action, in a way that science is not. Ethical theorising must tell us something about which actions are laudable, admirable or required. The laws of science, however, do not have to do with action so much as with the nature and interconnections of things.

At any rate, all the bicycle riding example shows is that knowledge of the laws of physics is not necessary to riding a bicycle. That is not at all the same thing as saying that physics could not plausibly motivate someone, if that person had the requisite knowledge, or that, by being motivated by such considerations, other important motives in that person’s life would be ruled out. Take another example, driving a car. It is obviously not necessary to understand the mechanics of engine combustion to know what to do when one is in the driver’s seat. We can admit that much without also being committed to the view that it is implausible that someone who had the relevant mechanical knowledge could be motivated by it, at least indirectly.68

Frank Jackson argues against objective consequentialism in much the same way. I end this sub-section with a lengthy quote from him, because I do not think the point can be more clearly put.

We need, if you like, a story from the inside of an agent to be part of any theory which is properly a theory in ethics, and (objective consequentialism) is a story from the outside. It is fine for a theory in physics to tell us about its central notions in a way which leaves it obscure how to move from those notions to action, for that passage can be left to something which is not physics; but the passage to action is the very business of ethics.69

1.4.2 External reasons and moral theories

We have seen that internal reasons are justified by external reasons, and we have also seen that one constraint upon external reasons is their capacity to be someone’s internal reasons. The question now arises as to the origin and justification of the external reasons themselves. Or, to put it another way, what grounds the independent ethical standard by virtue of which it makes sense to talk about external reasons? This, I suggest, is provided by moral theorising. If

---

68 Indirect motivation is discussed in the context of indirect consequentialism in § 2.2 and § 2.3.
69 Jackson, 1991, p. 467. Jackson is a subjective consequentialist. For further discussion on the difference between objective and subjective consequentialism, see § 2.4.1.
Jim is justified in killing the one because to do so brings about the best possible consequences, it is still open to us to seek a justification as to why consequences are what matter most in this case. We turn then to what consequentialist theorists have had to say on this question. External reasons are derived from and justified by moral theories. By a moral theory I mean a theory which tells us wherein rightful action consists and why. It follows that, if the relationship between internal and external reasons is as I have described it, the test of the adequacy of a moral theory is whether the justificatory reasons it provides could plausibly motivate. This test I refer to as the motivational test of moral theories.

Of course, this need not be the only test of a moral theory. Another test might relate to internal consistency. The claim here is not that providing a plausible set of motivating reasons is sufficient for a moral theory to be regarded as adequate; only that it is necessary. The motivational test of a moral theory is whether the theory could provide a plausible set of motivating reasons.

Some clarificatory points about the motivational test now need to be made. The first of these relates to the notion of plausibility. What is meant by a ‘plausible’ set of reasons?

One way of explicating this notion is to consider what the test would amount to if the notion of plausibility were omitted. If a theory had only to motivate per se, with no qualification, the test would be so weak as to be useless. After all, people perform all kinds of actions from all manner of motives. Motivational sets can be highly idiosyncratic. Even if we could not actually find a person who was always motivated by the considerations Kant presents in the *Groundwork*, such a person does not exceed the limits of our imagination. One way of strengthening the test would thus be to change ‘motivate’ to ‘motivate realistically.’ This would mean that a person so motivated would have to fall within the bounds of normality. The set of motives prescribed by the theory would have to be psychologically probable. This criterion, so amended, would rule out the possibility of someone being relevantly motivated, but at the cost of being overly eccentric or downright weird. This is an improvement on the first possibility, but it is not quite sufficient for our purposes. What is required is not just a life which is merely not-weird, but one which is morally admirable and personally satisfying. At the very least, the set of motives provided by a moral theory must not rule out the possibility of those goods which would normally be considered present in a good, meaningful and well-rounded life.

In Susan Wolf’s parlance, the question is, would a ‘healthy, well-rounded, richly developed character’ be ruled out if someone adopted the
set of moral motives prescribed by the theory and only that set of moral motives?\textsuperscript{70} We may refer to Stocker’s list of goods as a starting point. ‘Love, friendship, affection, fellow feeling, and community’ should not be excluded by the relevant theories.\textsuperscript{71} As we noted in the Introduction, whether these goods are properly considered to be moral or non-moral is not the issue. We can agree that they are important and should not be ruled out by acting as a moral theory prescribes. One problem with consequentialism and Kantian ethics – or so I shall argue in Chapters 2 and 3 – is that close personal attachments like friendship would be ruled out if someone were motivated by the reasons prescribed by these theories.

In deciding whether or not a motivational set is ‘plausible,’ we consider both the moral and non-moral motives which it rules out. How would we regard someone who did not have the motives which were ruled out by the theory? Would such a person be admirable in either a moral or relevant non-moral sense? In answering this question, we draw on pre-theoretical understandings of both morality and an admirable life. In this regard, the motivational test is not morally neutral; it requires us to make morally laden evaluative judgements. As I see it, this is no objection. Suppose we accept what was argued in the previous sub-section, namely, that moral theories should be able to prescribe action. It then seems difficult to see how moral theories can be evaluated without making some moral judgements about those actions which are prescribed, or about the life that someone would have if she acted accordingly. It was noted in the Introduction that the motivational test is a device for linking moral theory with practice. It expresses the thought that the plausibility of a moral theory is determined by looking at the actions it prescribes and the actions which, in the process, it rules out.

In applying the motivational test, our focus is not with the motives behind isolated actions or decisions. Our focus, rather, is with the sets of reasons given by moral theories. To put consequentialism to the motivational test, then, would be to ask whether the set of justificatory reasons provided by consequentialism could plausibly serve as a set of motivational reasons. To answer this question we look to whether consequentialism rules out motives which we could not plausibly wish to have ruled out.

The motivational test simply states that someone, somewhere must be capable of being motivated in the way prescribed by the theory. We do not have any particular person in mind, except that he or she must be capable of

\textsuperscript{70} Wolf, 1982, p. 419.
\textsuperscript{71} Stocker, 1987, p. 41.
meeting the ideals of character excellence outlined above. This hypothetical person is made deliberately vague, since we do not wish to narrow down the range of his or her possible motives unnecessarily. In making this fictional abstract character genderless, ageless and generally without identifiable dispositions, we leave open as full a range of motivational sets as possible.

The motivational test is to be distinguished from claims about motivational internalism. In the Introduction it was pointed out that the question of whether one should be an internalist or an externalist about motivation was to be side-stepped. We do not want the motivational test to be construed as a claim about motivational internalism. The claim about the motivational test is as follows.

(M) If it is right for any agent, A, to \( \phi \) for reason R, it must be plausible for some A to be motivated by R to \( \phi \).

The claim about motivational internalism is

(I) If A believes it is right to \( \phi \) then A will be motivated to \( \phi \).

The motivational test looks to the content of R, and asks whether R could count as a plausible motive. Internalism does not make any claims about the content of judgements or motives. Internalism is the claim that motivation is somehow ‘built into’ moral judgements, whatever those moral judgements happen to be. To put it another way, the motivational test is concerned with normative questions about motivation, whereas the internalist thesis is concerned with the meta-question of the connection between moral judgement and motivation.

We are able to accept the motivational test but still remain neutral on the question of whether we should be internalist or externalist about motivation. Suppose, for example, consequentialism were to pass the motivational test. Let us agree that this amounts to saying that promoting the best consequences could serve as a plausible motive of moral action. In agreeing to this, we establish only that someone could be so motivated without thereby having to surrender other motives which we would think integral to an admirable, desirable life. It would remain a separate question whether someone who judged an action to be right on consequentialist grounds would thereby be motivated to do it.

---

72 Frankena, 1958, p. 41.
A distinction needs to be made between primary and secondary motives. The distinction is drawn by Barbara Herman in the context of Kantian ethics but is relevant to understanding the motivational test. In summary, primary motives move the agent; secondary motives permit or rule out actions. For example, suppose a child wants to play at her neighbour’s house. She is motivated by the neighbour’s new computer game. She asks permission from her father. The father grants permission and the child goes. The child’s primary motive here is the prospect of the computer game. Her father’s permission merely enables her to proceed with what motivates her. The permission is a secondary motive. Secondary motives do not initiate action. Because primary motives give explanatory or internal reasons, they are the focus of the motivational test, rather than secondary motives. Moral theories must be able to motivate in the sense of giving primary motivation.

Primary motives are similar to Donald Davidson’s primary reasons. Davidson argues that ‘the primary reason for an action is its cause.’ We saw earlier how internal reasons serve two functions. Internal reasons answer the question, what is going to happen, according to S, and the question, what happened, according to S. That is, like Davidson’s primary reasons, internal reasons give causal explanations of actions. In the following three chapters a ‘primary motive’ is referred to in this causal sense. The secondary sense of motivation is not the kind of motivation which is relevant to internal reasons, and therefore is not the kind of motivation which is relevant to the motivational test.

The following three chapters put three moral theories to the motivational test. I argue that none of these theories can plausibly provide a set of motivating reasons.

---

73 Herman, 1993a.
74 Davidson, 1980, p. 4.
75 See § 1.3.1
76 The causal view of motives is also advanced by David Falk. A motive moves and impels in the sense that it ‘functions as a cause of action in the conventional sense of cause as an antecedent.’ (Falk, 1952, p. 495.) Understanding reasons in terms of causes is not uncontroversial. We shall see in § 3.1.2 that it is challenged by Kantians.
Conclusion

The distinction between internal and external moral reasons creates, on the face of it, a polarity between the considerations which motivate us in the daily world and the considerations prescribed by a pre-existing morality. Williams attempts to capture the middle ground between these two alternatives by emphasising the possibility of revising our motives through reflective deliberation. Reflective deliberation is supposed to bridge the gap between an agent’s transparent motives and the motives the agent would have if she were ideally rational or moral. External reasons are normative: they tell us what should motivate an agent according to some independent standard, and what would move the agent if she were ‘seeing matters aright.’ By making reflective deliberation as wide and open-ended as possible, the more of the ground putatively covered by external reasons is captured.

This attempt to claim the middle ground is not as successful as Williams would like. For an agent can reflectively deliberate only from within her particular motivational set. This motivational set may be flawed. In such cases deliberation does not provide the advantages of an independent standpoint. Williams can only provide a sufficiently critical standpoint by abandoning the idea of making reasons necessarily connected to individual motivational sets. He faces a dilemma. Either he retains the notion of reasons being confined to individual motivational sets, in which case he loses the benefits of having an adequate justificatory standpoint. Or he retains the notion of a justificatory standpoint, in which case internal reasons are no longer confined to individual motivational sets.

In respect of moral reasons, an alternative way of capturing the middle ground was postulated. The idea was to retain the notions of both internal and external reasons but show that they are not independent of each other. According to this view, internal reasons are criticised and justified by external reasons. And although external reasons are independent of any particular actual motivational set, they are not independent of what it is plausible for a motivational set to be. External reasons are limited by what it is plausible for people to hold as internal reasons. On this account, the Euthyphro question is answered by saying the gods desire those things which are good, but what can count as good has to be that which it is plausible for the gods to desire. Desires and external standards of desires are interdependent.

The contention that external reasons must plausibly be able to be internal reasons has implications for moral theories, since moral theories, in turn, justify external reasons. It follows that moral theories must be able to
provide plausible motivating reasons. Whether they can provide such reasons amounts to a test of the theory. In the following three chapters I put three theories to this test: consequentialism, Kantian deontology, and the ethic of care. I argue that none of these theories can provide a plausible set of motivating reasons. This is so because each theory rules out motives, either of a moral or non-moral kind, which could not desirably be omitted. Let us now turn to the first of these theories, consequentialism.
Reasons based on consequences

"Your majesty," resumed the Vizier, "there is an old proverb which says: "He who does not weigh the consequences of his acts shall never prosper.""

'The Fisherman and the Jinnee'
Tales from the thousand and one nights

According to consequentialism, the rightness of an action is determined by the value of the outcome or state of affairs which it brings about, or is expected to bring about. This idea has great appeal. For it seems compelling that we ought to prefer better states of affairs over worse states of affairs. Philippa Foot has remarked that this intuitively appealing idea haunts even those who are opposed to consequentialism.1 Similarly, the idea that one ought to be motivated by the thought of bringing about the best state of affairs has great appeal. Surely the problem, morally speaking, with many of our actions is that we give insufficient weight to their likely consequences. We frequently fail to take proper regard of the consequences of our actions on others, being motivated all too frequently by parochial self-interest. Would not the world be a better place if everyone, from individuals to heads of state to corporate leaders, were more motivated by the consequences of their actions, measured according to what is valuable? Haunting thoughts indeed for the critic of consequentialism.

In the previous chapter I argued that the test of the adequacy of a moral theory was whether it could provide plausible motivating reasons for action. At

---

1 Foot, 1988, p. 224. Samuel Scheffler calls the same idea "simple and seductive." (Scheffler, 1988a, p. 5.)
issue now is whether consequentialism can plausibly motivate. On many occasions it is desirable to be motivated by consequences. That much I have no wish, nor need, to deny. The claim of this chapter, however, is that it cannot always be the motive of moral action. Were we always to be motivated by consequences, other motives would be excluded which we would not wish to exclude. What are these motives? I shall argue that consequentialism cannot adequately motivate non-instrumental respect for persons; nor can it adequately motivate acts of friendship. Respect and friendship are missing from the consequentialist’s set of motives. Respect for persons, however, does fall within the ambit of Kantian ethics, and friendship within the ambit of ethics of care. These other two theories can pick up what consequentialism cannot accommodate.

The idea that consequentialism does not take account of the full range of human motivations, is a familiar one in the literature. It was noted in the Introduction that Bernard Williams’s critique of utilitarianism, and Michael Stocker’s critique of modern moral theories in general, both make use of the idea that consequentialism provides too narrow a set of motives. Such critics of consequentialism claim the theory is not psychologically realistic. They say that consequentialists, anxious to develop a systematic theory of what makes an action right, ‘ignore the basic facts of our moral psychology; they fail to keep in touch with the realities of motivation and deliberation, affection and reason, that are salient to ordinary subjects.’

There are three ways in which a consequentialist could respond. The first is to accept the motivational test and argue that consequentialism can meet it. Here the critic acknowledges that consequentialism must be compatible with normal, desirable motivations and argues that it is just that. The sophisticated consequentialist, argues Peter Railton, can overcome the problem of motivation by being only indirectly motivated by consequences. This appeal to indirect motivation, it is claimed, means that the consequentialist need not be constantly deliberating over options and perpetually calculating outcomes. Such a person’s motives arise from various habits and dispositions of a non-consequentialist kind, but such a one is only justified in having such dispositions because they are consequentially the best. This kind of response is commonly favoured by consequentialists and is the most challenging of the three.

---

3 Railton, 1984.
Another response is to deny that consequentialism has to concern itself with questions of motivation and human psychology at all. This second reply challenges the motivational test itself, denying that a moral theory needs to be able to provide motivating reasons. Proponents of this idea would say that consequentialism gives an account of right-making properties and, as such, it has no need to concern itself with human decision-making and motivation. Therefore, criticizing consequentialism because it cannot motivate is like criticizing a television because it does not bake bread: it was never intended to in the first place. The third response is to say that the motivational test already assumes consequentialism. The idea here is that all arguments against consequentialism which appeal to human psychology are consequentialist arguments anyway.

Modern consequentialist theories are descendants of the eighteenth and nineteenth century utilitarianism advanced by thinkers such as Jeremy Bentham, John Stuart Mill, John Austin and Henry Sidgwick. Indeed, much of the contemporary literature on the subject explores variations on themes and difficulties already discussed by these earlier thinkers. One of the difficulties in discussing whether consequentialism can plausibly motivate is specifying which version of consequentialism one is talking about. The literature is vast and contains many versions. In this chapter I refer mainly to Philip Pettit’s theory because, as one of the most well-developed accounts of consequentialism, it is able to accommodate, at least prima facie, many of the objections levelled at consequentialism. If my objections can stand against this account, they will go a long way towards standing against any account.

---

4 Bales, 1971.

5 For example, Frank Jackson’s argument defending consequentialism against the criticism that it cannot account for the ties we have to our ‘nearest and dearest’ had already been given by Austin and Sidgwick. (Jackson, 1991; Austin, 1996, Lecture 4, p. 163, and Sidgwick, 1930, p. 434. [Austin’s work first published 1832, and Sidgwick’s, 1874.]) More generally, the appeal to indirect consequentialism so familiar in contemporary discussions, had been advanced by Austin, Mill and Sidgwick. (Austin, 1996, Lectures 2 and 4; Mill, 1965, Ch. 2, pp. 269-271 [first published 1861]; 1987, p. 53 [first published 1884]; Sidgwick, 1930, p. 413.)

6 There is act-consequentialism (Singer, 1972; Smart, 1956, 1973), rule-consequentialism (Brandt, 1967; Gibbard, 1965; Toulmin, 1986), motive consequentialism (Adams, 1976; Parfit, 1984, pp. 25-26), indirect consequentialism (Hare, 1987; Jackson, 1991; McNaughton, 1988, pp. 177-180; Pettit and Brennan, 1986), satisficing consequentialism (Slote, 1984), objective consequentialism (Bales, 1971; Brink, 1989, Ch. 8; Railton, 1984), and subjective consequentialism (Jackson, 1991; Parfit, 1984, p. 25; Singer, 1977).

The chapter has four sections. In the first section I present an account of Pettit’s version of consequentialism, paying particular attention to the distinction he makes between ‘agent-relative’ value and ‘agent-neutral’ value (to be defined presently). In the second section I show how consequentialism cannot adequately account for a person being motivated by a respect for others. This is so, I argue, because respect is an agent-relative value and cannot be properly accommodated within the agent-neutral framework of consequentialism. In the third section I show how consequentialism cannot account for a further agent-relative value, that of friendship. In both the second and third sections I consider indirect consequentialism and argue that it cannot overcome the problem of motivation in the way that consequentialists would have us believe. The final section is devoted to the remaining two objections referred to above, namely that consequentialism need not motivate at all, and, secondly, that these arguments against consequentialism are themselves consequential.

2.1 Features of consequentialism

The basic idea behind consequentialism is that ‘there is one ultimate moral aim: that outcomes be as good as possible.’ There are, however, many variations on this basic idea. In ‘act consequentialism,’ the rightness of an action is determined by the value of the consequences which the action brings about, or is expected to bring about. Other types of consequentialism focus on other objects of valuation such as rules, dispositions or virtues. In rule consequentialism, for example, the right action is that which conforms to the rule which brings about the best consequences. What constitutes the best consequences is, in turn, determined by reference to some theory of the good or valuable. In utilitarianism, a common early form of consequentialism, the good is human happiness. For the act-utilitarian, therefore, the rightness or wrongness of an action is determined by the amount of happiness it brings about, or is expected to bring about, ‘for all mankind, or perhaps for all sentient beings.’ My focus in this chapter, however, is Pettit’s consequentialism.

2.1.1 Pettit’s consequentialism

Consequentialism, Pettit tells us, is the view that right action ‘is determined on the basis of the promotion of neutral – neutral and, of course, universal – values; it says nothing on what the relevant values are.’

Accordingly, Pettit’s consequentialism has two identifying features. The first has to do with the proper response to value and the second has to do with the kind of value which elicits the response. Consequentialism tells us that, whatever the values are, they must be neutral and they must be promoted. To promote a value, Pettit says, ‘is to maximise its expected realisation.’ It is to bring more of it about. As Stocker says, ‘whatever goodness is like, (what has) more of it is more obligatory than (what has) less of it.’

Promoting value is contrasted with honouring or instantiating a value. To honour a value, says Pettit, is to do your proper part under an idealised scheme of general compliance. He cites the following example to highlight the difference between promoting and honouring a value. Take the value of peace. As a pacifist, there are two ways in which one might respond to the value. One might act in such a way as to bring about the most peace. That may mean taking up arms so that at a later point there will be more peace overall. Or, alternatively, one might refuse ever to take up arms. In this latter case one is always peaceful in one’s actions even if, as a consequence, there will be less peace overall at a later point. In the former case, one promotes peace; in the latter, one honours peace. When one honours a value one focuses on one’s own actions at the present time; when one promotes a value one takes into account other people’s actions in some future foreseeable state of affairs.

The second feature of consequentialism relates to the kind of value which is to be promoted. Rightness is determined with respect to neutral values, says Pettit, not relativised values. The distinction between agent-relativity and agent-neutrality is a familiar one in the literature. It may be drawn in respect of values, rules, aims, theories or reasons. Pettit describes the difference between neutral and relative values as follows.

---

14 McNaughton and Rawling, 1992, p. 835. Two well-known accounts of the distinction in respect of values are given in Nagel, 1986, Ch. 9 and Parfit, 1984, p. 27 and §§ 55-58. See also McNaughton and Rawling, 1991. The distinction was already present in Nagel’s account of objective and subjective reasons in The possibility of altruism. (Nagel, 1970, pp. 91, 95.)
The value will be a neutral value, we can say, if and only if we can know what it is that is valued without knowing who the valuer is. It will be a relativized value, on the other hand – an agent-relative or agent-centred value ... – if and only if we cannot know what it is that is valued without knowing the valuer's identity.\footnote{Pettit, 1997, p. 125. See also Pettit, 1987, p. 75.}

An agent-neutral value is not indexed to a particular agent, whereas an agent-relative value is. It follows that an agent-neutral value can be maximised without any reference to who holds it, whereas we cannot know whether an agent-relative value is maximised without referring to the agent who holds the value. For example, take the value of truth-telling. If truth-telling is an agent-neutral value, it does not matter who tells the truth. What is important is that the truth gets told, and any given individual is just the means by which that happens. However, if it is an agent-relative value, it is indexed to a particular agent. If I hold truth-telling as an agent-relative value, what I value is my telling the truth.

2.1.2 The plausibility of being motivated by consequences

The question now is, how well can the promotion of neutral values fare as a motive? The claim for which I wish to argue is not that consequences (understood in terms of neutral values) can never plausibly motivate us. Rather, the claim is that is that the promotion of neutral value cannot be our only moral motive.

The remainder of this chapter is devoted to showing that consequentialism cannot capture the full set of our moral motives. However, in dwelling upon the limitations of consequentialism, we ought not to be blinded to its virtues. Being motivated by consequences is no bad thing much of the time. Pettit's point that consequentialism forces us to broaden our outlook is well made. Consequentialism exhorts us to see beyond what Pettit calls 'minding only your own moral business,' and requires that we see the overall effects of our actions.\footnote{Pettit, 1997, p. 168.} In being so motivated, we recognise our place in a broader social and moral scheme, and acknowledge our interconnectedness with others. The agent-neutral perspective acts as a great dampener on our powerful tendencies towards egotism. It remains true, nonetheless, that consequentialism
has motivational blind spots, and to two of these we shall shortly turn. But first it needs to be explained why the challenge to consequentialism is best understood in terms of the distinction between agent-relative and agent-neutral values.

2.1.3 Agent-neutral and agent-relative values

We have seen that consequentialism is distinguished by two features: the promotion of neutral value. How should a challenge to consequentialism best be presented? Should it query the promotional aspect or the agent-neutral aspect?

The first point to observe is that promoting goes hand-in-hand with neutral value and honouring aligns similarly with relative value. When rightness is understood in terms of honouring (as opposed to promoting) it involves acting on principle. In other words, when we honour the value of truth-telling we act upon the principle of truth-telling. Thus, if one honours loyalty and honesty, one acts upon principles which enshrine those values. Honouring in this sense refers to one’s ‘allegiance to what is right.’

We see, then, that to honour a promise is to pay special regard to one’s own actions in a certain way. One must act so that the promise is fulfilled. In honouring a value, one is not concerned with what others do but with what one does oneself in respect of the thing to be honoured. Because of this focus on oneself and one’s own actions, honouring is an agent-relative concept.

If honouring always aligns with agent-relative values, is it the case that promoting always aligns with agent-neutral values? Not quite. It still makes sense to speak of promoting agent-relative values. We can see that this is so by carefully considering the two propositions which Pettit says go to the crux of consequentialism.

1. Every prognosis for an option, every way the world may be as the result of a choice of option, has a value that is determined, though perhaps not up to uniqueness, by the valuable properties realized there ...
2. Every option, every possibility which an agent can realize or not, has its value fixed by the values of its prognoses: its value is a function of the values of its different prognoses, a function of the values associated with the different ways it may lead the world to be.\(^{20}\)

Let us suppose I have the agent-relative value of promise-keeping. This means I should act on the principle of promise-keeping in my own life. Now let us analyse the decision-making process for a particular case of promise-keeping. Let us say I have made a promise. I have two options in respect of the promise. The first is that I keep my promise. In doing so, a certain amount of promise-keeping value is realised. For the sake of argument, call this amount (1). The second possibility is that I do not keep my promise. No promise-keeping value is realised. The value is therefore (0). Which possibility is better, the first or the second? The first is better because (1) is a higher amount of value than (0). Having assigned a value to each option, I am now able to rank the two options in favour of the first.

In taking the first option I am therefore promoting relative value. In this process, both of Pettit’s propositions are realised. Firstly, a value can be assigned to each way the world may be as a result of a choice, in this case (1) and (0). Secondly, the value of each choice is a function of the assigned values. We can then rank the options.

If I have relative values I will rank options according to whether those relative values are realised. Therefore, in honouring a promise, the consequentialist may say that what I am really doing is promoting the agent-relative value of promise-keeping. Because of this, the distinction between a consequentialist and a non-consequentialist point of view is not best expressed in terms of the difference between *promoting* and *honouring*, but in terms of the difference between *promoting neutral value* and *promoting relative value*. It might be thought that the sense in which relative value is promoted is a trivial one. It amounts to saying that, on any given occasion, the difference between, for example, truth-telling as an agent-relative concept and truth-telling as an agent-neutral concept is that, in the latter case, I maximise truth-telling in general, whereas in the former case, I maximise my own truth-telling on this particular occasion. To speak of one’s own truth-telling as ‘maximising one’s truth-telling on this one occasion’ may seem to trivialise the meaning of

Reasons based on consequences

• 'maximise.' Nonetheless, this way of understanding what is at issue between the consequentialist and the non-consequentialist has its advantages. At the outset of this chapter I referred to Foot's remark that consequentialism captures a very intuitively appealing idea: that we ought to prefer better states of affairs over worse states of affairs. Understanding the debate in terms of agent-relative and agent-neutral value shifts the focus from states of affairs to kinds of values. Under the proposed way of understanding the issues at stake, both the consequentialist and the non-consequentialist can agree that it is the best state of affairs that counts. Where they disagree is over how the best state of affairs is to be determined. Is the best state of affairs one in which neutral value is promoted or one in which relative value is promoted? Is morality concerned with agent-neutral value alone, agent-relative value alone, or a combination of both? That is the issue at stake. Put this way, consequentialism loses much of its intuitive appeal.

A distinction drawn by Max Weber clarifies the difference between agent-neutral and agent-relative values. In his essay, 'The profession and vocation of politics,' Weber distinguishes between two different kinds of moral demands. In what he calls the ethics of conviction (Gesinnungsethik), our duty is to do certain things independently of how others respond to our action, or of how they intend to act of their own accord. We are responsible only for the actions which we perform. If bad consequences follow from an action performed out of pure conviction then, according to this way of understanding ethics, the doer is not responsible. 'The world ... or the stupidity of others, or the will of God who made them thus,' is responsible. In the ethics of responsibility (Verantwortungsethik), however, we 'make allowances' for the shortcomings of others. We understand ourselves as also being responsible for what others do as a result of our actions. Each individual holds herself responsible, not just for what she herself does, but also, to the extent that it is possible, for the way the world goes. As Thomas Nagel was to put it later, she holds herself responsible for what happens, not just what she does.

21 See Vallentyne 1988 for a discussion on this point.
24 Nagel, 1986, p. 165. The happening-doing distinction is another way in which consequentialism has been distinguished from non-consequentialism. See McNaughton and Rawling, 1991.
What distinguishes these two kinds of demands is the scope of responsibility which we believe is incumbent upon us to accept. Are we responsible only for acting upon our convictions (convictions which may well be grounded in the belief of their being universally applicable), or are we responsible for others' actions as well? Do we act as though it were a perfect world, in which all we have to do is our own small part, or do we acknowledge the imperfection of others and take that into account in our deliberations?

In following an ethic of responsibility, as consequentialists do, we see ourselves as responsible for the way the world goes. In doing so, we accept the imperfections of others, and take them into account in our calculations. In following an ethic of conviction, on the other hand, we take the view that we can only properly be responsible for our own actions. If we understand ethics as operating within a theological framework, the tension between an ethic of conviction and an ethic of responsibility evaporates. Here is how. Even though it may often seem that acting on our convictions (that is, agent-relative values) does not bring about the greatest overall welfare, this is only because our understanding is finite. Although we cannot grasp how agent-relative and agent-neutral values may be reconciled, we must have faith that they are. There is a 'divine convergence' between the two kinds of values. In those cases where acting upon our convictions does not bring about the best consequences we need not be too concerned. We can only be responsible for our own actions; we cannot be responsible for the way the world goes. That is God's responsibility and He has organised the world in such a way that the overall benefit of everyone is best served if each of us simply does what is right.

In the absence, however, of this theological framework, we may think that we should take on more responsibility for how the world goes. After all, if we mere mortals do not assume this responsibility, no-one will. This is the consequentialist's position. The consequentialist opts for an ethic of responsibility. He sees himself as responsible for the way the world goes, whereas one who opts for an ethic of conviction sees himself as responsible only for his own convictions. Convictions are agent-relative in kind.

---

25 The idea of divine convergence was put forward by Joseph Butler. See Larmore, 1987, p. 136 and Hare, 1987, pp. 213, 222. (Butler, of course, did not use the language of 'agent-relativity,' and 'agent-neutrality.') For more on Butler's 'divine convergence' see § 5.2.3(i).
27 According to this way of understanding the difference between consequentialism and non-consequentialism, the consequentialist takes the imperfections of others into account in his deliberations, whereas the non-consequentialist does not. Pettit implicitly acknowledges this point when he talks about the agent-relativist's ideal of others'
The distinction between agent-relativity and -neutrality helps to bring into sharp relief the differences between consequentialists and their critics. Opponents of consequentialism, although they may not use the language of agent-relativity and -neutrality, do take issue with consequentialism's agent-neutral flavour. Williams’s objection to utilitarianism, for example, does just that.28 Utilitarianism, he says, operates on the assumption that each individual is interchangeable with every other individual; it does not give preference to one over the other. Williams is here taking up John Rawls’s thought that utilitarianism ‘does not take seriously the plurality and distinctness of individuals.’29 In particular, says Williams, utilitarianism requires that the agent, in his decision-making, not assign any special significance to his own convictions. And yet, these convictions help give shape to a person’s life. Williams points out that people have commitments, projects and affiliations which give meaning and direction to their lives. Morality cannot demand of us that we abandon, in the name of overall utility, the very things that give our lives meaning.30

Interestingly, consequentialists themselves do not necessarily understand consequentialism as being inconsistent with the kind of agent-centred concerns brought to our attention by Williams. Pettit writes

Were consequentialism to reject the propriety of individual-centred concerns,
then it would be a highly revisionary and, to my eye, a very unattractive moral theory.31

To accommodate certain agent-relative values, consequentialists such as Pettit appeal to a theory of indirect motivation. Indirect motivation has a similar role in consequentialism to that played by God in the theological framework just compliance. He says that if you believe that you should honour rather than promote peace, then you believe that in this, the imperfect world, you should still do your part by the value. Even if others do not comply, that is not your concern. You stick to your ideal of their compliance. To honour a value, therefore, is to behave in a way that would promote it in a suitably complying context, even if it does not promote it in this, the actual, imperfect world. (Pettit, 1997, pp. 127-130.)

28 Williams, 1973a.
29 Rawls, 1986, p. 29. (First published 1971.)
30 Owen Flanagan also argues that any acceptable moral theory must accept what he calls ‘the personal point of view.’ By this he means that it is constitutive of being a person that one has a distinct point of view and distinct projects and commitments. Furthermore, these projects and commitments give each life whatever meaning it has. (Flanagan, 1991, Ch. 3. See also Scheffler, 1994, Ch. 3.)
31 Pettit, 1997, p. 96. As indicated in the Introduction, Jackson (1991) has a similar view.
outlined. That is to say, it putatively reconciles agent-relative values with agent-neutral values. The idea is that we are permitted to have agent-relative values but only because they ultimately promote agent-neutral values.

This relationship between agent-relative values and indirect consequentialism is explored in the following two sections, wherein I argue for the validity of agent-relative values on two grounds. The first relates to the idea that others place upon us certain obligations. These obligations arise from the obligation to respect others. Drawing on the work of Stanley Benn, I shall argue that consequentialism cannot adequately account for the moral notion of respecting persons. The second ground on which I shall argue for the validity of agent-relative values relates to the way we treat those who are personally close to us. I consider the value of friendship and argue that to value friendship properly is to value it in an agent-relative way. Just as respect for others is an agent-relative value, so too is friendship. The consequentialist’s reply will be to argue that it is justifiable to be motivated by relative value only because such motives end up promoting neutral value.

So far we have not heard an argument in favour of the promotion of neutral values over relative values. If we accept that there are these two kinds of values, on what grounds should neutral values be preferred over relative values and why should relative values have to be justified only through reference to neutral values? What is so objectionable about relative values? Before moving on to defend the two kinds of relative values just mentioned, I should like to consider Pettit’s argument against relative value.

2.1.4 Agent-relative value and universalizability

Pettit’s argument against agent-relative value in favour of agent-neutral value is that only the promotion of neutral value can guarantee ‘the universalizability constraint on judgements of rightness.’ Pettit believes that valuer-relativist theories fare so badly in respect of this constraint that ‘we should give up on them in favour of a consequentialist approach.’

There are two ways we might defend valuer-relativist theories. One is to deny that universalizability is so important. The other is to deny that valuer-relativism means we have to give up on universalizability. As it happens, I do not think universalizability is quite as tight a constraint as Pettit makes out, but I do not wish to press that point here. Arguing against the

---

33 Pettit, 1997, p. 142.
universalizability of moral judgements is a large undertaking, and is not necessary for present purposes.\textsuperscript{34} Rather, I take issue with Pettit’s claim that agent-relative value necessarily fares so badly in respect of universalizability.

To clear the ground for Pettit’s argument let us imagine two people, John and Mary. Each of them believes in keeping his or her own promises. In the language of consequentialism, they both hold the agent-relative value of promise-keeping. However, they hold it in different ways. John believes promise-keeping is a value for everyone. Not only should he keep his promises, but his wife, his children, his political masters and anyone else too (at least in relevantly similar circumstances).

Consider now Mary’s way of holding the value. She believes that promise-keeping is something she should do as part of her personal code of ethics. However, she makes no claim about what others should do. Whether or not they keep their promises is their own affair and a matter for their consciences. She does not think that the values which bind her necessarily bind others. Both John and Mary take responsibility only for their own actions in respect of keeping promises. They do not take responsibility for the way the world goes in respect of promise-keeping. Because the scope of their responsibility is confined only to their own actions, rather than to maximising the amount of value in the world, they both have an agent-relative value.

If holding an agent-relative value meant doing it Mary’s way, there would be a problem for universalizability. Holding a relative value and acting upon it in this way would amount to each person going off and doing her own thing, so to speak, according to whatever values she happened to be attracted to. It could very easily result in everyone having quite different values from one another and no means of justifying those values. Pettit’s point about agent-relative values undermining universalizability would then be well made. However, the concept of agent-relative values does not entail Mary’s way of acting. In acting on one’s convictions, one is by no means necessarily acting on values that are idiosyncratic to oneself. Agent-relativity does not mean agent-idiosyncrasy.

Let us return to John’s way of holding the agent-relative value. John’s method rather than Mary’s is Pettit’s target when he says that agent-relative value is not universalizable. First of all, we need to be clear about what is meant by universalizability. It means, says Pettit, that

\textsuperscript{34} Others have taken up the challenge. See Dancy, 1993; Noddings, 1984; Potter and Timmons, 1985; Wiggins, 1987a and 1987b and Winch, 1972, Ch. 8 (and for a reply to Winch, Kolenda, 1975). For proponents of the universalizability test see Hare, 1955 and 1963; Mackie, 1985.
if we say that an agent \( A \) ought to choose option \( O \) in circumstances \( C \) — these may include the character of the agent, the behaviour of others, the sorts of consequences on offer, and the like — then we assume that something similar would hold for any similarly placed agent.\(^{35}\)

He says that it is clear how consequentialism can meet the challenge of universalizability. If option \( O \) is the right option, it is because of the neutral values which \( O \) promotes: ‘If those neutral values make \( O \) the right option for me in those circumstances, so they will make it the right option for any other agent in such circumstances.’\(^{36}\) Pettit believes that non-consequentialism, however, cannot meet this test of universalizability. His argument may be summarised as follows.\(^{37}\)

1. If \( O \) is universalizable, then in judging it is right for me to do \( O \) in circumstances \( C \), I commit myself to there being a normative reason for me and for any arbitrary agent \( X \) to prefer doing \( O \) in \( C \).
2. In ‘perverse’ situations the best way for me to satisfy the preference that \( X \)’s in general do \( O \) is to choose non-\( O \) myself.\(^{38}\)
3. But the agent-relativist (non-consequentialist) will choose \( O \) in such perverse situations.
4. Therefore, from (2) and (3), in perverse situations I do not satisfy the preference that \( X \) does \( O \).
5. Therefore, from (4) and (1), in perverse situations an agent-relative judgement is not universalizable.

What is wrong with this argument? We need to keep our eye on premise (2) because this premise sneaks a consequentialist assumption into the argument. The agent-relativist can agree that there are, indeed, situations where the best way of ensuring a general pattern of \( O \) in the world at large is for the agent to choose non-\( O \). This is simply another way of saying that there are times when the best way of maximising \( O \) in the world at large is to do non-\( O \). Once we introduce the goal of maximising what people in general do, the

---

35 Pettit, 2000a, p. 179.
36 Pettit, 2000a, p. 179.
37 This argument is repeated throughout the paper (Pettit, 2000a), but is best gleaned from the discussion on p. 181.
38 Examples of such perverse situations would be where one takes up arms to secure peace at a later point, or where ‘the best way to get people to help their children (is) to proselytize and not pay due attention to one’s own.’ (Pettit, 2000a, p. 181.) Scheffler says such situations have ‘an air of paradox.’ (Scheffler, 1994, p. 127, 1994a, pp. 133-145.)
consequentialist has won the day. But why should the agent-relativist concede that ensuring a general pattern of $O$, or maximising $O$ in the world at large, should be her goal? Or, to put it another way, why should universalizability be taken to entail maximisation of neutral value?

Slippage has occurred between premise (1) and premise (2). By premise (1) we are required to believe that any arbitrary agent $X$ should do $O$. But in premise (2) we are required to ensure that as many $X$s as possible do $O$. We need a connecting premise between (1) and (2) along the lines of: ‘If I prefer that $X$ do $O$, I must prefer that as many $X$s as possible do $O.’ Call this premise (1*). To see what is wrong with the sequence of (1), (1*) and (2), let us consider the example of keeping promises. If it is right for me to keep a promise (in circumstances $C$), then by the universalizability principle (1), it is right for you to keep your promise. Furthermore, by (1) and (1*), if it is right for me to keep a promise then it is right for me to bring it about that as many people as possible keep their promises. This is the same as saying that if it is right for me to keep my promise then I must maximise promise-keeping. Therefore, even before we get to premise (2), the consequentialist has assumed (by premise (1*)) that the rightness of my keeping my promise entails the maximisation of promise-keeping. But this is the very point at issue. To be sure, the consequentialist will want to make this assumption, because once this assumption is granted, the rest of the argument will go the consequentialist’s way. But making this assumption begs the question in favour of the consequentialist. We have already seen that the very issue of disagreement between the consequentialist and the non-consequentialist is over how to interpret the rightness of promise-keeping. Whereas the consequentialist says, ‘if promise-keeping is right then I must maximise the total amount of promise-keeping,’ the non-consequentialist says, ‘if promise-keeping is right then I must keep my promises.’

Premise (1*), being the culprit premise, deserves closer examination. It says that if one has a preference for promise-keeping then one has a preference for as much promise-keeping as possible. The difference between the agent-neutralist and the agent-relativist is that, in the former case, the preference is overriding in all situations, whereas it is not overriding for the agent-relativist. While it is true that the agent-relativist has a preference that as many people as possible perform the relevant kind of action, $O$, she is not prepared to bring this state of affairs about at any price. In particular, she is not prepared to bring it about in those cases where she must do not-$O$. In fact, what makes her an agent-relativist is precisely that other people’s actions at a future point do not carry as much weight with her as her own actions at the time of decision.
Let us now see how the agent-neutralist and agent-relativist fare in respect of universalizability if we omit the question-begging \((1^*)\). By \((1)\) we may say that, ‘if it is right for me to keep my promise in circumstances \(C\), then it is right for anyone else to keep their promise in those circumstances. So far so good. Now let us suppose that \(C\) is the perverse situation. That is, the circumstances are such that I have to choose between Option (i), keeping my own promise, and Option (ii), maximising the promise-keeping of others by not keeping my own promise. (Suppose Option (ii) enables five other promises to be kept.) Let us consider each of these options.

For Option (i), premise \((1)\) now reads, ‘if it is right for me to keep my own promise in the perverse situation, then it is right for anyone to keep their own promise in the perverse situation.’

For Option (ii), premise \((1)\) now reads, ‘if it is right for me to bring it about that five promises are kept (by not keeping my own promise), then it is right for anyone to bring it about that five promises are kept (by not keeping his or her own promise) in the perverse situation.’

The universalizability principle does not tell us whether we should choose Option (i) or Option (ii). It only tells us that, whichever option we choose to be the right one, it follows that it would also be the right option for any other agent similarly situated.

Therefore, the agent-relativist can still universalize her judgements. It is just that the way in which she universalizes her judgements will not satisfy the consequentialist. For what is being universalized from the agent-relativist’s point of view is the judgement that all arbitrary agents in \(C\) should act upon their convictions, rather than the judgement that all arbitrary agents in \(C\) should maximise neutral value.

Pettit’s argument only serves to restate for us the difference between the agent-relativist and the agent-neutralist. The difference is that the agent-neutralist will, in perverse cases, maximise the relevant value, whereas the agent-relativist will not. But Pettit has not succeeded in showing that the agent-relativist must give up on universalizability. He has only succeeded in showing us that the agent-relativist must give up on maximising neutral values. But \(that\) is no news: we always knew that the agent-relativist had to give up on maximising neutral values. What defines the agent-relativist is precisely that she does not understand herself as being responsible for maximising other people’s actions; she only understands herself as being responsible for her own actions.
It should be emphasised that this argument against Pettit does not establish that one should always be an agent-relativist; it does not show that one should never maximise agent-neutral value. We have simply shown that, in those cases where someone is motivated by agent-relative values, it is not thereby entailed that she give up on the universalizability of her judgements. We have now shown that Pettit's only argument in favour of promoting agent-neutral values over agent-relative values is unsustainable.

2.2 Consequentialism and respect for others

Having clarified the issues at stake between the consequentialist and the non-consequentialist points of view, the question now is, can consequentialism plausibly provide motivating reasons for action? In light of the discussion in the previous section, this question could be put another way: how well can neutral values fare as motives? In this and the following section I argue that there are some agent-relative values which motivate in their own right, and not because they are based on a commitment to certain other neutral values. These agent-relative values are: respect for persons, and friendship.

Let us now anticipate the consequentialist's response to the challenge from agent-relativists. There are two ways for the consequentialist to go. One way is to deny that there are any legitimate agent-relative values. The other would be to agree that there are such values, but argue that they are only justified because they promote agent-neutral values. It would be difficult for the consequentialist to deny the first possibility. Clearly, people do have agent-relative values. We care not only about what happens in the world, but about our own actions as well. Great religions are based upon agent-relative values. The Ten Commandments, for example, give a list of relative values since they are concerned with what people should and should not do themselves, rather than what should or should not happen in the world at large. Certain kinds of actions are prohibited, – killing, adultery, idolatry – whatever the consequences. Going down the second path appears more fruitful and the consequentialist will reply to the arguments in the following two sections by taking such a route. This second possibility brings us to indirect
consequentialism. The idea is that people may be motivated by agent-relative values but are only justified in doing so because such actions ultimately promote neutral values. Relative values, therefore, are a means of promoting neutral values. Let us now examine indirect consequentialism in more detail.

2.2.1 Indirect consequentialism

Consequentialists have long realised that neutral values could not always plausibly motivate, at least, not in an immediate, direct way. Nineteenth century utilitarians like Mill and Austin vehemently denied that utilitarianism required the principle of utility to be the motive for all action.\(^{41}\) ‘The principle of utility,’ says Austin, would be ‘a halting and purblind guide’ to action.\(^{42}\) It would be ‘a gross and flagrant error’ to think that one’s conduct at the moment of action should be guided by utility calculations.\(^{43}\) Rather, by cultivating the proper habits of thought, and by a suitable education, one develops the proper sentiments towards various classes of action. According to Austin, the principle of general utility guides the education of the sentiments but is not itself a direct motive.\(^{44}\) Speaking of the utilitarian, Austin says that

\begin{quote}

though he approves of love because it accords with his principle, he is far from maintaining that the general good ought to be the motive of the lover. It was never contended or conceited by a sound, orthodox utilitarian, that the lover should kiss his mistress with an eye to the common weal.\(^{45}\)
\end{quote}

Sidgwick also thinks that the utilitarian need not always be motivated by the principle of utility. He says

\begin{quote}

... the doctrine that Universal Happiness is the ultimate standard must not be understood to imply that Universal Benevolence is the only right or always the best motive of action. For ... it is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim: and if experience shows that the general happiness will be more satisfactorily attained if men frequently act from other motives than pure universal philanthropy, it is
\end{quote}

\(^{41}\) See n. 5, this chapter.
\(^{42}\) Austin, 1996, Lecture 4, p. 116.
\(^{43}\) Austin, 1996, Lecture 4, p. 120.
\(^{44}\) Austin, 1996, Lecture 2, pp. 119-120.
\(^{45}\) Austin, 1996, Lecture 4, p. 163.
obvious that those other motives are reasonably to be preferred on Utilitarian principles.\textsuperscript{46}

Consequentialists are well aware that there would be problems if consequentialism were taken to entail that people should be motivated by consequences in a direct and transparent way. After all, we are not commonly motivated on a day-to-day level by neutral values such as the overall happiness of humanity. Indirect consequentialism provides an answer to this problem. It contends that one can be a consequentialist without consequentialist deliberations being in the forefront of one’s motivational psychology. In other words, the consequentialist need only be \textit{indirectly} motivated by consequences. ‘Restricted’ consequentialism and ‘virtual’ consequentialism are versions of this same idea.\textsuperscript{47}

In her regular daily behaviour, the sophisticated, indirect consequentialist acts upon a variety of dispositions and values. She does not perpetually make consequentialist calculations and deliberate over each and every action. However, from time to time she may reflect on whether her dispositions are justified or whether they should be modified in some way. It is only at this point that she deliberates consequentially. She is only consequentialist ‘off-line,’ so to speak, in moments of reflection, when she distances herself from the hurly-burly of habitual motivations.\textsuperscript{48} For the most part, her consequentialism is a background belief, which does not consciously motivate. ‘There are many values whose promotion requires precisely that the agent not think in terms of them,’ says Pettit. In such cases, people

should put themselves on – better, leave themselves on – the more or less automatic pilot, associated with that mode of thinking, secure in the knowledge that that represents a habit of mind which is justifiable, indeed uniquely justifiable, in the consequentialist currency to which they pay allegiance.\textsuperscript{49}

\textsuperscript{46} Sidgwick, 1930, p. 413. Mill wrote something similar. Happiness, he said, should be the end of all actions or rules of action. However, this means that it is the justification of ends, ‘not itself the sole end.’ (Mill, 1987, p. 53. [First published 1843.])

\textsuperscript{47} Pettit and Brennan, 1986 and Pettit, 1997. Railton’s ‘sophisticated’ consequentialism is also a variation of indirect consequentialism, as we shall see in § 2.4.1. (Railton, 1984.)

\textsuperscript{48} This is Pettit’s account (1997, pp. 156-162.) Richard Hare has a similar view based on a distinction between intuitive and critical thinking, where the latter takes place in a ‘cool hour.’ (Hare, 1987, p. 227.)

\textsuperscript{49} Pettit, 1997, p. 158.
Indirect consequentialism permits people to retain their agent-relative values. Austin’s lover may still direct his attention exclusively to his mistress, but is only justified in doing so because such close personal attachments ultimately work for the common good.

It is worthwhile pointing out an ambiguity in the notion of indirect consequentialism. It is not always clear whether the indirect consequentialist means to say that consequences still motivate, but at a level removed from spontaneous daily responses, or whether such a one means to say that consequences do not need to motivate at all. In the latter case the claim would be that consequentialism is a theory about right-making properties and has nothing to do with decision procedures, even at an indirect level. In this section by ‘indirect consequentialism,’ I mean the first of these options. In the following section, I consider the latter possibility.

The consequentialist has much to gain from this idea of indirect motivation. Not only does it permit, at least at first blush, agent-relative values to be accommodated. It also overcomes another objection which might be raised against consequentialism, namely, that being a consequentialist requires that a person be excessively calculating in her decision-making. Perpetually calculating expected utility values would not be an efficient decision procedure. It takes too much time and effort and in many cases it cannot be done at all, because one cannot always predict accurately what the results of one’s actions will be. Many situations call for quick responses. Acting upon well-cultivated habits and sensibilities is more likely to produce the right outcome in many situations because it does not require deliberation. Furthermore, preoccupation with calculating, deliberating and weighing up options would diminish qualities of personal excellence such as spontaneity and naturalness of demeanour. Such personal qualities may well be among the goods to be promoted. The sophisticated consequentialist has no need to continuously calculate because she habitually has the right responses and outlook.

2.2.2 Two kinds of indirect consequentialism

The question now is, does indirect motivation work? That is, does it enable consequentialism to pass the motivational test? The answer is no – at least, not

---

50 There is a further problem, of a more theoretical kind, namely that such continuous calculating leads to a regress. If we regard deliberation as an action, which it seems reasonable to do, we must first deliberate over how much time to allocate to the deliberation. But then we must deliberate over how much time to allocate to the second deliberation. And so it goes *ad infinitum*. (See Railton, 1984, pp. 153-154.)
in all cases. Not all agent-relative values will comfortably fit within the indirect consequentialist’s framework. To understand why, we need to examine more closely how indirect motivation works. A distinction was made in the previous chapter between primary and secondary motivation.\footnote{See § 1.4.2.} A primary motive moves a person to action. A secondary motive acts like a traffic light, either permitting or refusing the action to take place. Here is the problem. When consequentialism motivates indirectly, it must do so as either a primary motive or a secondary motive. If it motivates secondarily, it does not pass the test. This is so because the kind of motivation required by the test is primary motivation. That is, consequentialism must be able to move people to action, not just rule actions in or out. If, on the other hand, consequentialism motivates primarily, then it is subject to the same criticisms as direct consequentialism, at least in some cases. Let us now consider each of these possibilities.

(i) Indirect consequentialism and secondary motivation

First of all, how could consequences figure as a \textit{direct} secondary motive? What makes a motive direct is that it has immediate force. What makes it secondary is that it either permits or prohibits actions. Suppose we want to perform a certain action. Before doing so, we run a quick mental check of the likely consequences of the action with respect to various values. The action we want to perform has disvalue. Because of this we refrain from doing it. What makes this a case of \textit{direct} motivation is that the calculation occurs just as we are about to take action.

Consequences could work as an \textit{indirect} secondary motive when we reflect upon habits or dispositions which we have already acquired or are considering cultivating. From time to time we might reflect on the various dispositions and habits which we have acquired through a process of cultural and familial socialisation. If it turns out that the habits promote the relevant neutral values, our habits are given a green light. On the other hand, if, on reflection, we realise that they do not promote neutral values but are instead perhaps self-indulgent, they are given a red light. Such habits are unjustified and we should resolve to undo them.

The motivational test that we set up in the previous chapter was that a justificatory reason must plausibly be able to be someone’s explanatory reason. Explanatory reasons must be able to tell us what moves a person to action. Secondary motives, as we have seen, cannot do this. They permit or prohibit
actions or habits, but do not initiate them. Therefore, for the purposes of the motivational test, secondary motivation is not sufficient. Consequentialism must be able to provide primary motives.

(ii) Indirect consequentialism and primary motivation

The other way in which consequences can indirectly motivate is by providing primary motivation. This is how it would work. I wish to promote neutral values. To do so I see that I need to cultivate certain habitual responses. I set about cultivating those responses. In this case we develop the habits because they provide the most effective means of attaining the desired outcomes. When consequences work as a secondary motive they act as a gatekeeper on existing motives. When they work as a primary motive they impel us in a certain direction.

Here is an example of indirect primary motivation. Imagine a tightrope walker aiming to keep his balance. He discovers that the best way of keeping his balance is by not thinking about it. He decides that each time he walks the rope he will think about roses instead. In this example his balance is the value to be promoted. However, he is only indirectly motivated by considerations of keeping his balance. He is directly motivated to think of roses; he is indirectly motivated to keep his balance. So too the indirect consequentialist has to forget about consequences (on some occasions) just so that the best consequences can be brought about.

There are two points to note about the tightrope walker. These may seem obvious and unremarkable, but they will become crucial later on to an understanding of why, in some cases, indirect consequentialism is disanalogous to the tightrope walker. The first is that the goal of keeping his balance has motivating force. It motivates him to put thoughts of balancing out of his mind. The second point is that thinking of roses is just a means to an end. He does not want to think about roses for its own sake, but because it happens to help him achieve a certain end, namely, his balance. If, through a quirk of his psychology, it happened that the best way of achieving this end was for him to focus on keeping his balance at all times, then that is what he would do.

Let us now try to summarise all these distinctions. Consequences figure as an indirect secondary motive when one reflects upon one’s habits and dispositions with a view to their justifiability. When consequences are an indirect primary motive, one deliberately sets about trying to develop the habits and practices which will bring about the best consequences. The secondary motive regulates our habits and actions, while the primary motive
Reasons based on consequences

initiates our habits and actions. What makes the motive direct or indirect is the point at which the deliberation occurs. If it occurs immediately and consciously before the action, then it is a direct motive. If it occurs, however, ‘off-line’ and irregularly, then it is an indirect motive.

Having clarified what indirect consequentialism is, I shall now argue that certain agent-relative values cannot be properly accommodated by it.

2.2.3 Respect as an agent-relative value

To show that consequentialism cannot provide a complete set of motivating reasons, we must show that at least some agent-relative values are valid. We have already established that the most plausible response to this suggestion on the part of the consequentialist is to argue that there are agent-relative values but such values are always justified by agent-neutral values. In such a case the agent-relative value becomes a means of promoting the agent-neutral value. However, whether an agent-relative value can properly be understood as simply a means of promoting neutral value will depend upon the kind of agent-relative value it is. I will consider first a simple example of an agent-relative value where its subordination to an agent-neutral value is plausible, before moving on to a more complex example where such subordination is implausible.

Consider Martha, who holds and acts upon the neutral value of recycling. She promotes recycling in her workplace and through her efforts in her local conservation group. She becomes a public spokesperson for the group, advocating better recycling plants and recycling practices in her community. She is so busy promoting recycling that she has no time to organise recycling in her own home. She hurriedly tosses all her domestic waste into her garbage bin, whether it is recyclable or not. This is the very practice she is so committed to changing in others. What would we say of Martha? Most likely, we would frown upon Martha’s poor personal recycling credentials. If recycling is what others should do, why does she not do it herself? Should she not practise what she preaches? We often believe that someone who has a discrepancy between what they promote publicly and what they practise privately loses moral authority. Such inconsistencies give rise to allegations of hypocrisy.

Knowing that such hypocrisy, or at least the perception of such hypocrisy, would adversely affect her promotion of recycling, let us suppose that Martha decides that she should adopt what looks like a more agent-relative approach to recycling. She resolves to adopt a domestic recycling program: everything which can be recycled, she recycles. In doing so, she avoids embarrassing allegations of hypocrisy and thus improves her moral authority.
She is also now better placed to promote her cause by setting an example. By becoming the very model of the modern recycling woman, she is in a position to inspire her neighbours and others with whom she comes into contact.

In this case, Martha’s agent-relative reforms are motivated by consequentialist considerations. Agent-relativity is thus sucked up into the consequentialist vacuum cleaner (to use a metaphor of David McNaughton and Piers Rawling52). But all is not quite as it seems. If we revert to the language of convictions it will become clear why. We recall that to act on agent-relative values is to act on one’s convictions. If Martha recycles only to avoid allegations of hypocrisy, to establish her moral authority and to inspire others, then she is not really acting on her convictions. She is creating the appearance of having convictions for a certain effect. For her really to be acting on convictions she has to recycle just for the sake of her convictions about recycling: she must recycle just because recycling is the right thing to do.

Relevant here is a distinction which will reappear in the following chapter, between conforming to duty and acting for the sake of duty.53 One conforms to duty when one’s action coincides with what duty requires, but the motive is something other than duty, such as expedience. One acts for the sake of duty when the action is done purely from the motive of duty; duty is the reason for the action. The story we have told about Martha parallels the idea of an action conforming to duty, but not being done for the sake of it. Thus, in reforming her recycling ways, Martha acts in conformity with her convictions. However, because her actions are not performed for the sake of that about which she has convictions, it is not a genuine case of acting on conviction. If it is done for consequentialist reasons it is the semblance of acting upon convictions, a ‘going through the motions’ to have a certain effect. Mimicry of the act is not the same as the act itself.

Moreover, interestingly, to the extent that mimicry has value, it only has so because the genuine act has value. Therefore, we only value the appearance of acting on principle because acting on principle is itself so respected and valued. In Machiavelli’s The Prince, politics is portrayed as the art of appearances. The Prince must appear to be honest, loyal and candid without actually being so. The idea is to gain the advantages of being virtuous without acquiring the vulnerability that goes with virtue.54 But why would one wish for the outward manifestations if the actual thing itself were not valued? The value of appearances shadows the value of the real item itself. The good

53 See § 3.1.1.
54 Machiavelli, 1952, p. 93. (First published 1532.)
Reasons based on consequences

consequences of appearing virtuous only come about because people value genuine virtue. Yet, in acting virtuously to get the good consequences one is no longer genuinely virtuous. What is true in Machiavelli’s case for virtue is also true for convictions.55

The point here is that, if taking a stand on principle is done from consequentialist motives, that is, to create a certain effect or impression, then one is not really taking a stand on principle at all. In other words, to act upon agent-relative values as a means of promoting agent-neutral values is self-defeating. The agent-relative character of the values is destroyed once the motivation becomes agent-neutral.

The consequentialist’s riposte, however, is easy to anticipate. Such a one might concede our recycling example. The consequentialist could agree that, if Martha recycles not from her convictions about recycling, but from her desire to impress others, then she is not really acting upon her convictions. But it still remains open to the consequentialist to ask, why does she have that conviction? The consequentialist can quite plausibly claim that she has it because recycling brings about good consequences. That is to say, the agent-relative value of recycling is only justified on consequentialist grounds. In reasoning this way, the consequentialist moves to a higher order level of justification. This is the indirect consequentialism which was discussed in the previous sub-section. The story runs like this. Martha has a genuine conviction that she should recycle. This means that she believes that she herself should recycle. But why does she have this conviction in the first place? Because recycling reduces waste and helps solve environmental problems. This is a consequentialist justification. The agent-relative value of recycling is ultimately justified by the neutral value of waste reduction. In this case Martha is directly motivated by her convictions about the rightness of recycling, but she is indirectly motivated by consequences.

The question now is, are all agent-neutral values, all convictions, able to be justified in this way? Or, to put it another way, will the move to indirect consequentialism work in all cases? No. It works in the recycling case only because recycling is an instrumental (agent-relative) value. We do not value recycling in and of itself; we value it because it brings about other desirable

55 It is a moot point whether it is morally better to give the outward appearance of virtue, whilst actually remaining self-seeking or vicious, or whether it is a better thing not to bother even to keep up the appearances. At least if one keeps up the appearances, and makes oneself something of a hypocrite, one is tacitly acknowledging the virtue. As la Rochefoucauld eloquently put it, ‘L’hypocrisie est un hommage que le vice rend à la vertu.’ (Hypocrisy is the homage which vice pays to virtue.)
states of affairs, such as reducing landfill, greenhouse gases and the like. For this reason a plausible case can be made for a consequentialist justification for recycling. But not all agent-relative values are instrumental in this way. In the remainder of this section I show how the value of respect for persons, as an agent-relative value, does not fit within the indirect consequentialist’s framework in the same way that the recycling example does.

Let us now consider the debate between Benn and Pettit over whether consequentialism can motivate respect for persons.\textsuperscript{56} In his book, \textit{A theory of freedom}, Benn distinguishes between reasons of concern and reasons of respect, the former being what he calls ‘value-centred’ and the latter being ‘person-centred.’ Reasons of concern look to the good consequences of an action and whether the action will preserve, sustain or promote a valued and valuable activity, object or state of affairs. These reasons ‘require that we optimise – that we make the world as good as it can be made.’\textsuperscript{57} Person-centred reasons, on the other hand, according to Benn, involve considerations relating to principles such as freedom, justice, rights and truthfulness. All these principles relate to our dealings with other people, and we are justifiably committed to them simply because the subject is a person. The difference, then, between value- and person-centred reasons is that the former directs our attention to a property which should be promoted, whereas the latter directs us to an action we should perform in respect of a person. It is clear that value-centred reasons align with Pettit’s neutral values and person-centred reasons align with agent-relative values.

Benn says that respect for others, as a person-centred reason, is a ‘basic deontological notion.’\textsuperscript{58} In his 1989 paper, ‘Consequentialism and respect for persons,’ Pettit takes issue with Benn’s argument that only deontology can adequately account for the idea of respecting persons. Pettit wants to show that one can respect others and still be a good consequentialist. There are two possible strategies open to him. One would be to deny that respect for persons is an agent-relative value or, in Benn’s terminology, that it constitutes a person-centred reason. Here the idea would be to argue that respect for persons is best understood as a neutral value. What would it mean for it to be a neutral value rather than a relative value? Holding respect as a neutral value would mean that we should try to bring about as much respect as possible in the world, even if that means showing disrespect to some. The second strategy is similar to that presented in the recycling case. The idea here would be to show that respect is

\textsuperscript{56} Benn, 1976; 1988 and Pettit, 1989.
\textsuperscript{57} Benn, 1988, p. 8.
\textsuperscript{58} Benn, 1988, p. 240.
Reasons based on consequences

a relative value but that it is only justified on consequentialist grounds. It is this second option that Pettit chooses. In doing so, he opts for an indirect consequentialist approach to the value of respect.

Now, to the crux of the matter. Pettit says that ‘what is really moving Benn in his claim that a consequentialist cannot endorse reasons of respect is the phenomenology of deliberation.’ 59 In support of this view, he cites Benn as saying

a person-centred reason affects deliberation quite unlike a value-centred reason. …
A person-centred reason can impose a constraint on an otherwise appropriate action without there being any alteration in the evaluation of expected outcomes. 60

Let us be quite clear about the claim Pettit says that Benn is making. It concerns a comparison between person-centred reasons and value-centred reasons. Person-centred reasons, says Pettit, do not involve the same kind of deliberation with respect to outcomes as value-centred reasons. By this Pettit means that there is no weighing and evaluating of outcomes in the way that there is with value-centred reasons. Value-centred reasons involve more deliberation because they are explicitly concerned with outcomes. Showing respect for persons does not require any such deliberation because it is not concerned with maximising the outcome. 61

Pettit takes Benn’s reference to deliberation as being a reference to the phenomenology of deliberation, that is, a reference to how people actually experience deliberation. We may summarise what Pettit takes Benn’s argument to be in the following way.

(1) All value-centred reasons involve deliberation.
(2) Respecting persons does not involve deliberation.
(3) Therefore, respecting persons is not a value-centred reason.

61 In a later paper Pettit advances a similar interpretation of non-consequentialists in general. He says non-consequentialists ‘focus on deliberation rather than justification and, noticing that it will often be counter-productive to deliberate about the promotion of a value involved in action – a value like loyalty or respect – conclude that in such cases choices are justified by honouring the values, not by promoting them.’ (Pettit, 1991, p. 239.) Although he refers here to responses to value rather than kinds of values, his point – that non-consequentialists focus on deliberation – remains the same.
We recall that Pettit wants to argue that value-centred reasons can accommodate respect for persons. He argues against Benn by challenging (1). Pettit says that value-centred reasons do not necessarily involve deliberation. The consequentialist, so he argues, is not committed to any claims about phenomenology. That is, one can be a consequentialist and not have to enter into deliberative calculations. The indirect consequentialist, as we have seen, does not have to enter into consequentialist calculations. Therefore, according to Pettit, one can respect persons if one is an indirect consequentialist.

Pettit poses the question of whether there is 'any property such that it is a plausible value for a consequentialist to want promoted and such that the best way of promoting it is, plausibly, for agents to act on person-centred reasons.' In other words, he is asking whether there could be any consequentialist justification for acting upon person-centred reasons. Such a justification would require a value which was promoted by respecting persons. Yes there is such a value, he says, and the value is 'a certain sort of interpersonal security' prevailing in people's dealings with one another. We should respect each other and do so in a way which is not subject to the vagaries of consequentialist calculations because that enables us to feel secure with each other and trust each other. Thus, according to Pettit, the consequentialist respects persons because to do so enables conditions of trust to arise. Trust is necessary for there to be social relations at all.

This, in a nutshell, is the consequentialist story: as a consequentialist one promotes neutral value. One such value is that of social trust and stability. Question: how may this value best be promoted? Answer: by showing respect for others, that is, by acknowledging their various rights. Respect for others is thus perfectly compatible with consequentialist reasoning. We should note that such a consequentialist does not have to be constantly deliberating about whether they should show respect. Social trust is best promoted where respect is automatic and not subject to revision at every turn. The consequentialist thus cultivates the necessary dispositions to act in a respectful way, spontaneously and non-deliberatively.

What is wrong with this attempt by the consequentialist to accommodate person-centred reasons by appealing to indirect motivation? We recall that Pettit supposes Benn's point to be that consequentialism cannot account for person-centred reasons because such reasons are non-deliberative. Pettit therefore thinks that all he needs to do is show how the consequentialist

---

can be non-deliberative and he has overthrown Benn's contention that consequentialism cannot account for respecting persons. However, it is a mistake to think that Benn is appealing to, or must appeal to, the phenomenology of deliberation. Person-centred reasons are incompatible with consequentialism, not because they are non-deliberative, but because of their connection to moral personality. More specifically, persons demand respect just because they are persons and not because such respect has certain spin-offs, such as bringing about trust and security. On the consequentialist account there is no reason to respect persons other than that it happens to bring about security and trust. The consequentialist makes the value of respect instrumental in nature: if it turned out that disrespecting people brought about interpersonal security, then disrespect we would. But there is no reason, in principle, why some other value might not serve this purpose even better, thus rendering respect unnecessary. Benn understands respect as having a stronger command over us than Pettit's account permits. We do not need to look to trust or social stability, however important they may be, to provide a reason to respect others. According to Benn, respect for persons contains its own justification.

To make Benn's claim about the incompatibility of consequentialism and respect for persons sound plausible we need to consider the nature of person-centred reasons. When we show respect for persons, says Benn, we recognise their moral personality. Moral personality is, in turn, derived from natural personality. Natural personality refers to a person's capacity to make choices, to act upon reasons, as opposed to just being swept along by circumstances outside of her control. However, it is not just agency which is necessary for moral personality; one must be conscious of one's agency. A natural person is able to choose actions on the basis of her beliefs and has an awareness of that capacity.

To be a natural person is to possess, and to be aware of oneself as possessing, certain causal capacities. It is to distinguish oneself from the things in the world which are simply the subjects of happenings, carried along by the tide of events.\(^{64}\)

It is the agent's intentionality which makes the difference between happenings and actions. The agent is able to form intentions in the light of preferences, principles and commitments. Such a person "knows himself as thinking and

\(^{64}\) Benn, 1988, pp. 90-91.
feeling ... (and is) aware of himself as a decision-maker or chooser.' Not every human being meets this criterion. According to Benn, infants and some mentally ill people would not fall into this category.

The transition from natural personality to moral personality is somewhat elusive. According to Benn, when we understand ourselves as natural persons we develop a concept of ourselves as moral persons too, that is, as entitled to a degree of forbearance from others. In this way, natural personality, that is, our capacity for agency, gives rise to moral personality. Having moral personality means that we can legitimately claim respect from others, and must ourselves show respect to others. ‘Claiming respect – the recognition of our moral personality – on the grounds of our natural personality, we are then committed to extending it to anyone else satisfying the same conditions.’ When we respect others, we acknowledge their agency. We acknowledge that they are intentional agents, capable of acting on their own reasons and choosing their own projects and commitments.

According to Benn, reasons of respect have their origin in the recognition of others’ agency. Others are owed forbearance and consideration because, like ourselves, they are project-makers. Another’s project claims our respect, not because it is valuable and worthy of our concern, or even because we have an interest or benevolent concern for its author, but simply because it is a person’s project. In respecting another’s project, which was freely chosen by that other person, we show respect for that person himself. Benn’s account of respect, with its focus on the moral significance of agency and intentionality, is strikingly Kantian in outlook.

Benn says that recognising the moral personality of others, that is, showing respect, makes trust possible between people. Not only that, but ‘For a culture that made the connection (between natural and moral personality), it would be irrational to abandon it at the cost of losing these social advantages.’ This may seem to amount to a consequentialist justification of respect. But the distinction between primary and secondary motivation is critical here. Benn is appealing to consequences as a secondary motive: we have reason not to abandon respect because such abandonment would have bad social consequences. This is quite a different matter from

---

65 Benn, 1988, pp. 92-93.
66 Benn, 1988, p. 94.
67 Benn, 1988, p. 98.
68 Benn, 1988, p. 107.
69 The connection between friendship and a Kantian approach to ethics is discussed in §§ 3.2-3.4.
70 Benn, 1988, p. 101.
saying that the reason we respect in the first place is the good social consequences.

The important point about actions motivated by respect is that they contain their own justification, at least on Benn’s account. They are not justified simply on the grounds that they promote certain neutral values. This means that actions which show respect for others cannot in principle be substituted by any other actions for the sake of achieving a better outcome in terms of neutral value. In other words, in respecting another person, one is not aiming (either directly or indirectly) to maximise some other neutral value. This is why, when acting upon a person-centred reason, one does not have to enter into calculations about how much neutral value is being maximised. Pettit supposes that because a story can be told about how one can respect people without deliberating about it, that therefore consequentialism can accommodate person-centred reasons. But this objection misses the point, since it is not the absence of deliberation, in and of itself, which is important.

The move to indirect consequentialism — as a way of showing that consequentialism can account for respect for persons — is unsatisfactory. In the previous section we discussed the tightrope walker who is indirectly motivated by the desire to keep his balance. We established that, although thoughts of keeping his balance are not foremost in his mind, they nonetheless have motivating power. It follows that indirect consequentialism does not do away with consequences as a motive; it simply does away with them as an immediate motive. But on Benn’s account of respect for persons, respect is not motivated by consequences, either directly or indirectly. Respect is motivated by its connection to moral personality. Respect is not just a socially useful attitude, it is morally required of us in its own right.

Benn’s Kantian-based account of respect shows us exactly what is so dubious about an agent-neutral justification of respect. Again, the agent-neutral approach makes respect contingent on some greater agent-neutral outcome, in this case, the promotion of trust. And again, in doing so, it denies the intrinsic value of agency and autonomy. The advantage of Benn’s agent-relative account of respect is that it captures the intuitive thought that others are to be respected for their own sakes, not as a means to some further end, even if that end is a socially benevolent one. As it stands, Benn’s account of person-centred reasons is undefeated.
2.3 Consequentialism and friendship

In recent years a number of authors have suggested that consequentialism alienates us from those with whom we have close personal attachments. In this section I take up this idea, arguing that friendship, like respect, is an agent-relative value which cannot be properly accommodated within the consequentialist framework. The motives necessary for friendship cannot be reconciled with the motives required by consequentialism.

Again, there are two ways the consequentialist can respond. The first is to dispute that friendship is best understood as being an agent-relative value. The second is to concede that it can be understood as an agent-relative value but that to do so is only justifiable on agent-neutral grounds. This second response is the appeal to indirect consequentialism. I consider each of these possibilities in turn.

2.3.1 Friendship as an agent-relative value

Before considering the question of whether friendship is best understood as an agent-relative value or an agent-neutral value, a few remarks about features of friendship are in order.

One of the difficulties of trying to say anything philosophically significant about friendship is the enormous variety of relationships which go by the appellation, ‘friendship.’ Friendship comes in many flavours and degrees. When to call someone a friend varies not only from person to person but from culture to culture (hence the tendency of some cultures to regard Australians as shallow because of the readiness of the latter to call someone they have just met at a summer barbecue a ‘friend’). For the purpose of discussing friendship in the context of consequentialism, I want to draw on a distinction made by Aristotle between valuing a friend instrumentally and valuing a friend for his or her own sake.

We value people and our relations with them for various reasons. In Anne Brontë’s novel, Agnes Grey, the heroine, Agnes, converses with Rosalie about the forthcoming marriage of Agnes’s sister, Mary. Rosalie wants to know about Mary’s fiance. Is he rich? Is he handsome? What sort of house does he have? Upon learning that he is neither handsome, nor rich, nor does he have a large house, Rosalie is disgusted. Agnes replies ‘You did not ask me if

Mr Richardson were a good, wise, or amiable man; I could have answered yes to all these questions.\textsuperscript{72} Rosalie and Agnes have different ways of valuing a marriage partner. The former is only interested in the material benefits which the gentleman in question is able to bestow upon Mary, whereas Agnes is interested in the man himself, his \textit{character}. This distinction, between the benefits a person can give to another, and the person himself, lies at the heart of Aristotle's distinction between different kinds of friendship.

Aristotle distinguishes between friendships which are valued because they are useful and those which are valued for their own sake. The former are instrumentally valuable: they help us obtain a goal beyond the friendship, such as financial gain, social connection, career advancement and power. These friendships are contingent on our purposes and on the capacity of the friend to fulfil those purposes. Consequently, says Aristotle, they 'are easily dissolved, when the friends do not remain similar to what they were; for if someone is no longer pleasant or useful, the other stops loving him.'\textsuperscript{73}

In what Aristotle calls ‘complete friendship,’ by contrast, the friend is loved for no other reason than for who he or she is. There is no end beyond the friendship other than the friendship itself. Michel de Montaigne captures this idea in his essay on friendship:

\begin{quote}
If I were pressed to say why I love him, I feel that my only reply could be: ‘Because it was him, because it was I.’\textsuperscript{74}
\end{quote}

In complete friendship each party wishes goods for each other for the sake of the other. Such friendships may be advantageous and pleasurable but the advantages and pleasures are outcomes rather than motives for the friendship. Only virtuous people, says Aristotle, have the capacity for such friendships. The more virtuous a person is, the more he or she is capable of complete friendship. It so happens that in \textit{Agnes Grey} we see this theory borne out. Agnes’s and Rosalie’s different attitudes towards Mary’s fiance are indicative of the differences in their characters. Agnes, who is a portrait of virtue, has lasting, flourishing friendships, while Rosalie, with her interest in worldly goods and appearances, only has relationships which are useful to her.\textsuperscript{75}

\begin{itemize}
\item \textsuperscript{72} Brontë, 1988, p. 132. (First published 1847.)
\item \textsuperscript{73} NE, 1156a, 20.
\item \textsuperscript{74} Montaigne, 1988, p. 97. (First published 1580.)
\item \textsuperscript{75} As a consequence of her preoccupation with the external goods of human relationships, Rosalie finds great unhappiness in a materially advantageous but loveless marriage. For Aristotle there would be a necessary connection between her lack of virtue, her incapacity for complete friendship and her personal unhappiness.
\end{itemize}
Similarly, Montaigne distinguishes between friendships which are valued instrumentally and those which are not. Like Aristotle, he values the latter more highly.

All those relationships that are created and fostered by pleasure and profit, by public or private interests, are so much the less fine and noble, and so much the less friendships, in so far as they mix some cause, or aim, or advantage with friendship, other than friendship itself.  

This is one of the hazards of power and wealth: it is difficult to know who one's friends really are. As Orson Welles once remarked, 'When you are down and out something always turns up -- and it's usually the noses of your friends.'

But more needs to be said about what it means to like someone 'for her own sake?' The idea here is that we attend and respond to our friend's character -- who she is -- rather than the benefits which will accrue from our association with her. We do not have our own pleasure or benefits in mind, nor the benefits which may accrue to the universe in general. Of course, it is also true that we may have difficulty sustaining a friendship which was unpleasant and a drain on us and from which we gained no advantage. But that is not to say that those considerations are what motivate us in cultivating and nurturing the friendship. We seek out friends just so that we can have friends.

Neera Badhwar Kapur suggests that one key difference between instrumental and complete friendships is replaceability. In complete friendships, says Kapur, the friend is loved for her essential rather than incidental features.

These include both her character traits -- the fundamental intellectual, psychological, moral and aesthetic qualities that constitute an individual's personality -- and her unique perspective on herself and other: her view of the important and unimportant, her interest in herself and others.

We should note, however, that it is not always easy to tell which are the essential features of another, and which are the incidental features. William Newton-Smith gives the example of 'a dynamic capitalist entrepreneur whose

---

76 Montaigne, 1988, p. 92.
personality is intimately bound up with the acquisition of wealth.' Bankruptcy for such a person may well bring about a personality change.\textsuperscript{80} With the occurrence of such a personality change it becomes difficult for the friend to know whether she is responding to the change of fortune or the change of character.

We may grant that the distinction between essential and incidental features may be blurred. Nonetheless, the point that friendship can be its own reward remains true. To explain why one seeks friends, cultivates friendship, acts out of commitment to friendship, no further justification other than the friendship itself is required. To have relationships distinguished by mutual goodwill, affection and pleasure is a good in itself.\textsuperscript{81} The ensuing discussion is concerned with friendship of this non-instrumental kind.

Are we to understand friendship as an agent-neutral or agent-relative value? That is the question. Pettit claims that friendship, along with values such as happiness and freedom, is a neutral value.\textsuperscript{82} Imagine, he says, I decide that what is important in life is ‘the enjoyment of personal loyalties, whether the loyalties of family or friendship.’ There is a question about how I should respond to this value.

Should I honour the value in my own life, devoting myself to developing the bonds of kith and kin? Or should I only permit myself such devotion so far as that is part of the more general project of promoting the enjoyment of personal loyalties? Should I be prepared to use my time in the manner most effective for that project even if the cost of doing so – say, the cost of spending so much time on journalism and politics – is that my own personal loyalties are put under severe strain?\textsuperscript{83}

His answer is quite clear: ‘The agent should honour the values only so far as honouring them is part of promoting them, or is necessary in order to

\textsuperscript{80} Newton-Smith, 1973, p. 123.

\textsuperscript{81} We need to distinguish between two ways in which pleasure and friendship may be connected. In the first instance pleasure is the end which is sought and friendship is simply a means of obtaining it. Aristotle regards such friendships as incomplete, in the same way that friendships motivated by utility are incomplete. The goal of pleasure renders the friendship instrumentally valuable. (NE, 1156a-1156b.) Alternatively, pleasure may be understood as being internal to the friendship; the pleasures arise out of the affectionate relationship which is sought and sustained for its own sake. (For a discussion on the intrinsic value of friendship see Tom Regan, 1983, esp. p. 145.)

\textsuperscript{82} Pettit, 1997, p. 124.

\textsuperscript{83} Pettit, 1991, p. 231.
promote them. Pettit’s response here, although expressed in the language of honouring and promoting, is that we should understand friendship as a neutral value. This means that if we value friendship, we should promote the total amount of friendship even if it means in the process we miss out on having any friends ourselves.

But this way of valuing friendship does not conform to our common understanding of what valuing friendship amounts to. When we observe that someone values friendship we do not mean to say that they go around promoting it and that they see their own friendships as just another way of promoting friendship in general. To take an extreme case, the person who campaigned against birth control in the name of promoting friendship – if there were a larger population, there would be more friendships – would have a most peculiar way of valuing friendship. Indeed, it is hard to imagine such a person. Friendship is valued in an agent-relative way. Having an agent-relative value of friendship here is taken simply to mean ‘being a friend.’ Thomas Scanlon makes the point – about how we value friendship – in terms of reasons for acting.

A person who values friendship will take herself to have reasons, first and foremost, to do those things that are involved in being a good friend: to be loyal, to be concerned with her friends’ interests, to try to stay in touch, to spend time with her friends, and so on. He goes on to say that one who values friendship will also believe she has reasons to cultivate new friendships and will also think it good for other people to have friends. These latter two reasons, which have to do with the maximisation of friendship, show how friendship may be understood as an agent-neutral value, while the first set of reasons (given in the quote above) refer to it as an agent-relative value.

Scanlon is perfectly correct to understand that the value of friendship is ‘first and foremost’ to be understood in terms of agent-relativity, rather than agent-neutrality. But the consequentialist’s reply is readily anticipated. Such a one will claim that Scanlon has made the error of unnecessarily conflating two distinct things: valuing friendship and being a good friend. Whether these two things are entirely distinct is, of course, precisely the point at issue. In what follows I should like to defend Scanlon’s idea that our commitment to

---

friendship is primarily agent-relative. This is not to rule out friendship as an agent-neutral value. It is just to say that the valuing of friendship in a purely agent-neutral way cannot plausibly be morally required of us. Indeed, there is something peculiar, if not ridiculous, in valuing friendship in this way.

Consider Miss Muddle, who, convinced of the virtues of consequentialism, and inspired by a recent lecture by a prominent philosopher of that persuasion, Professor Pettifogger, sets about promoting friendship. Being an efficient and single-minded woman (as well as one who takes her moral obligations seriously), she establishes the Society for the Promotion of Friendship in next to no time. The singular thing about the Society for the Promotion of Friendship is that none of its members has any friends. They have no time for friends; they are much too busy promoting friendship. Miss Muddle is so busy organising the Society’s gatherings, raising funds, keeping the books and recruiting members that she neglects and loses her old friends and has no time for new ones. When an old friend, Rose Worthy (from the wanton, pre-Society days), contacts her, Miss Muddle quickly dismisses her. No, she has no time to talk. Nor has she the interest in doing so. Her time is precious and she is well aware that she can promote more friendship in the world by efficiently organising the AGM for her Society than by allocating the same amount of time to being friendly with Rose. ‘Sorry, Rose, I’m too busy,’ she sharply announces. She has no qualms in spurning Rose because, after all, in devoting herself to the promotion of neutral value in this way, she is fulfilling her moral obligations.

What can we say about Miss Muddle? One response is that there is a deep irony in her life. There she is, so busy promoting other people’s friendships that she has no time for her own. If we were in Rose’s position, we might feel a stronger reaction than irony. We might feel Miss Muddle was hypocritical. Rose might say (perhaps to a friend who is not a member of the Society) ‘That Madge Muddle is so hypocritical. She’s always going on about the importance of friendship, but she’s incapable of being a friend herself.’ As we saw earlier in the context of the recycling example, we regard someone as a hypocrite when there is a discrepancy between the views they espouse publicly and their own actions. Wilcox observes quite rightly that it would be ‘distressing to discover that the best moral conception of what we ought to be like ... might suggest that we should at least closely resemble hypocrites.’86 Indeed, it would be more than distressing; it would render the theory untenable. Even if we did not wish to go so far as to call Miss Muddle a hypocrite, we

86 Wilcox, 1987, p. 84.
would certainly think there was something odd about the way she had arranged her life. Surely she has missed the whole point of friendship if she ends up not having any friends herself. This is so because friendship is the kind of value that, if we wanted to confirm whether someone had the value, we would look at that person’s own life.

There is a way in which we could understand Miss Muddle’s life as not being so odd. Suppose she had a sense of regret that she was unable to sustain the friendships she once had. Suppose she felt a loss. Her regret at the way things had turned out would show that she did still value friendship in the agent-relative sense, even though she was not herself a good friend. As it is, Miss Muddle, as a good consequentialist, has no such attitude. In dismissing Rose, she simply understands herself to have worked out her proper priorities. She values friendship in the agent-neutral sense alone.

Consider, on the other hand, the person who has the value of friendship in an agent-relative way. Rose Worthy values friendship in just this way. She is not interested in maximising friendships but in having and enjoying them. Like Aristotle and Montaigne before her, she values not the number of friends she has, but the quality of the friendships she does have.\(^{87}\) She makes time for them. She can be depended upon to help them in times of trouble and to celebrate with them in times of triumph. She confides in them and takes pleasure in their mutual society. What would we say about such a person? Most likely, we would observe how well adjusted she was. We may wish that such a person would include us among her friends.

If Rose declined to join the Society for the Promotion of Friendship, there would be nothing ironic or hypocritical in her decision. Those appellations are used to signify a gap between what one tells others they have an obligation to do, and what one does oneself. When one has an agent-relative commitment to certain values but does not take it on oneself to publicly promote them, there is no corresponding charge of hypocrisy. Of course, there may be other judgements made. We might think of such a one (particularly if we are consequentialists) as too concerned with her own integrity, too self-indulgent perhaps.\(^{88}\) Even so, in respect of our example, Rose could hardly be called morally self-indulgent just because she thinks she has more reason to care about her own friendships, rather than neglect them so she can devote herself to

---

\(^{87}\) Aristotle warns against seeking many friends. ‘Indeed,’ he says, ‘it even seems impossible to be an extremely close friend to many people.’ Friendships require a good deal of time and attention. (NE, 1170b, 20-1171a, 20.) Montaigne regarded one ‘perfect friendship’ as sufficient. (Montaigne, 1988, pp. 100-101.)

\(^{88}\) On the charge of moral self-indulgence see Williams, 1981c.
Miss Muddle’s Society. In caring about her own friends, she is simply being normal.

The irony which is so evident on the part of someone who only promotes a value such as friendship and kinship, even at the expense of her own bonds of friendship and kinship, makes such a person ripe for satire. Charles Dickens’s character, Mrs Jellyby, in Bleak House is just such a character. Mrs Jellyby was so busy promoting the welfare of the African natives of Borrioboola-Gha that she completely neglected her own children. Mrs Jellyby, who ‘had very good hair, but was too much occupied with her African duties to brush it,’ spent her entire day ‘in correspondence with public bodies, and with private individuals’ on matters relating to her African interests. Her eyes, it seemed, ‘could see nothing nearer than Africa.’ She had no time for the ‘frivolous’ tasks of child rearing. Her children, dirty ‘little unfortunates,’ ‘tumble about,’ unfed. The house is in chaos. Her promoting the welfare of African children at the expense of her own makes her a figure of mockery and mirth, not admiration.

Suppose Miss Muddle were to reorganise her priorities. After careful reflection she decides to rectify the inconsistency in her life. She decides to act out, or, as Pettit would say, to instantiate, the very values which she espouses to others. Now she is both a good friend and President of her Society. She holds friendship to be both a relative and a neutral value. Far from being critical of her practices, we now admire her. I take the examples of Miss Muddle, Rose and Mrs Jellyby to show that friendship is primarily an agent-relative value. Going all out to maximise friendship would only be laudable when the person can be shown actually to have it in their own life first (or at the very least to want to have it in their own life and have regret if they do not). In deciding whether someone has the value, we look first and foremost to their own actions. As Scanlon says, the reasons involved in being a good friend ‘take priority over the reasons we have to promote friendship.’ It is left open to us

---

89 Dickens, 1985, Chs 4 and 5. (First published 1853.)

90 Mrs Jellyby is said to be based upon the historical figure, Caroline Chisholm. (Kaplan, 1988, p. 287.) The chapter in which she first appears, ‘Telescopic philanthropy,’ was intended by Dickens to be a satire of the radical utilitarians of the day. Although appalled by the effects on the poor of industrial and social laissez-faire policies, Dickens was critical of the radical utilitarians. He referred to the ‘supernatural dreariness’ of the reformers associated with Bentham, and James and John Stuart Mill. In Dickens’s opinion they treated the imagination as a poor cousin of reason, logic and science. (Kaplan, 1988, p. 305.) John Stuart Mill, in turn, referred to Dickens as ‘That creature Dickens.’ (Stafford, 2000.)

to want to increase the total amount of friendship in the world but this is not required by the value of friendship.

We saw earlier that friendship may be sought for its own sake. If we understood friendship as being something to promote, as an agent-neutral value, rather than something to enjoy in our own lives, we would have to see our friends as expendable. If we saw an opportunity for two friendships to be made if we destroyed only one friendship, then we would be obliged to destroy the one friendship. Yet a person who thinks in this way would be incapable of friendship. Her friends would be there just to be picked up and cast aside, depending upon opportunities for maximisation. This would require us to see our friends instrumentally, as a means of maximising friendship in general. But to be a friend one needs to be capable of appreciating another just for her own sake, just for her own company, just for how one finds her. No, we do not prove how much we value friendship by our willingness to maximise it in the world at large. We show how much we value friendship by our willingness to maximise it in our own lives and by our willingness and capacity to cultivate the appreciative and interactive qualities required by friendship. None of this is to say that we might not deliberately end friendships, but, if we value friendship as an agent-relative value, the termination will be for reasons internal to the friendship. Perhaps, for example, the friend betrays us. Or perhaps we develop new interests and grow apart.

2.3.2 Indirect consequentialism and friendship

It still remains open to the consequentialists to appeal to indirect motivation, as they did in the case of the agent-relative value of respect. Such a line of response would go as follows. Friendship is an agent-relative value. People should go about having friendly relations in their own life and not worry too much about maximising the total amount of friendship in the world. However, it is only justifiable to act upon this agent-relative value because to do so promotes happiness overall. That is, the agent-relative value of friendship is justified because it promotes a certain neutral value.

This approach is subject to the same problems that arose in connection with the value of respect. In the case of indirect motivation, consequences must work as either a primary or secondary motive. To pass the motivational test they must work as a primary motive. For consequences to work as a primary indirect motive we would have to reason as follows. The stability of the community (or some similar value to do with happiness, trust, etc.) is a neutral value which, as a consequentialist, I wish to promote. The best way of
promoting such stability is if each person pursues friendly relationships. It is best for these relationships if they are pursued in a spontaneous, uncalculating way. Therefore, I will cultivate in myself a disposition to form such relationships in an uncalculating way.

As we observed in the discussion about the agent-relative value of respect, indirect consequentialism does not do away with the motivational problem altogether. It merely shifts the difficulties onto another level. It simply means that, rather than being motivated by considerations of consequences in an immediate, daily way, one is motivated by consequences at irregular points of reflection, when one's thoughts are 'off-line.' The point remains, however, that one is still motivated by consequences. One pursues the friendship because one can see that to do so promotes some relevant neutral value.

But this way of understanding friendship is not compatible with the nature of friendship. We have already observed that we cultivate friendship because we have a capacity to like others, to be attracted by their personality and mien and this is as it should be. We pursue friendships just to have friendships, not to serve some greater impartial good. We do not cultivate friendship because of the consequences it brings in a more general abstract sense. Friendship is its own end, not a means to some other end. In this respect, friendship is quite different from many other kinds of relationships. Compare friendship to neighbourliness. In the case of neighbourliness, one may have a more general goal such as neighbourhood stability, low crime rates, and so on. One may deliberately cultivate benevolent relationships with one's neighbours in the name of this greater goal. In the case of friendship there is no greater goal beyond the friendship itself.

Friendship, like respect for persons, is an agent-relative value of a non-instrumental kind. To subordinate the value to a greater neutral value is to distort the kind of value that it is. Yet, subordinating friendship and respect to greater neutral values is the only way in which consequentialism can accommodate these very important agent-relative values.

2.4 Two consequentialist replies

At the outset of the chapter I claimed that if a test of the worth of a moral theory was whether it could provide plausible motivating reasons, then consequentialism would fail it. We saw in the previous two sections how the consequentialist could attempt to deny that this was so. The indirect consequentialist claims that consequentialism can indeed provide motivating
reasons, but only indirectly. I argued that this move to indirect motivation did not enable consequentialism to pass the motivational test.

There are two further replies now open to the consequentialist. The first is that it is a mistake to suppose that consequentialism need motivate at all. That is, goes the objection, consequentialism is not a theory about how to make the right decision, but about what makes a decision the right one. The second reply is that the motivational test must itself appeal to consequences. These replies are the subject of this final section of the chapter.

2.4.1 Truth-conditions versus acceptance-conditions

Consequentialism, it might be said, is not a theory about how people should make decisions about which action or course of action to take; it is a theory about which actions are the best. If it is true that the purpose of consequentialism is not to provide a decision-making procedure, then consequentialism would not need to concern itself with questions of motivation. Eugene Bales is one author who has maintained the need for a sharp distinction between right-making properties and decision-making procedures; Railton is another.92

The distinction between right-making properties and decision-making procedures underpins Railton’s distinction between objective and subjective consequentialism.93 Objective consequentialism is the view that ‘the criterion of the rightness of an act or course of action is whether it in fact would most promote the good of those acts available to the agent’.94 Subjective consequentialism, on the other hand, prescribes a mode of deliberation. ‘Whenever one faces a choice of actions,’ says Railton, ‘one should attempt to determine which act of those available would most promote the good, and should then try to act accordingly.’95 A feature of objective consequentialism is that it ‘concerns the outcomes actually brought about’.96 On the other hand, because subjective consequentialism is concerned with guiding decision-making, it must refer to anticipated or expected outcomes, not actual outcomes. After all, when making a decision one can only take into account what one expects

92 Bales, 1971 and Railton, 1984, § 6. See also Brink, 1989, Ch. 8, § 11.
94 Railton, 1984, p. 152.
95 Railton, 1984, p. 152.
96 Railton, 1984, p. 152.
will happen. One does not know the actual consequences until after the event.\textsuperscript{97} Railton advocates objective consequentialism because it ‘has the virtue of not blurring the distinction between the \textit{truth-conditions} of an ethical theory and its \textit{acceptance-conditions} in particular contexts.’\textsuperscript{98}

We need to consider objective consequentialism more closely since it maintains that consequentialism need not concern itself with questions of motivation. There are two ways of understanding objective consequentialism. There is a moderate and an extreme version. Both are problematic.

\textit{(i) Moderate objective consequentialism}

The moderate version of objective consequentialism may be understood as being the theory propounded by Railton. Railton denies that objective consequentialism prescribes a decision-making procedure. That would be too crude. Instead, Railton tries to tell us how objective consequentialism would work in practice. ‘A \textit{sophisticated consequentialist} is someone who has a standing commitment to leading an objectively consequentialist life, but who need not set special stock in any particular form of decision making and therefore does not necessarily seek to lead a subjectively consequentialist life.’\textsuperscript{99} Such a one does not bring a consequentialist calculus to bear on every act.

Contra Railton, however, it is not true to say that this version of objective consequentialism has nothing to say at all on the question of whether one should be motivated by consequences. In effect Railton’s sophisticated consequentialist is an indirect consequentialist. To be sure, the sophisticated consequentialist is not motivated in a direct and clumsy way before each and every act. But he must be \textit{indirectly} motivated. After all, by Railton’s own admission, the sophisticated consequentialist has a ‘standing commitment’ to consequentialism. What could this mean other than that he is indirectly motivated by consequences? To have a commitment to a criterion of rightness is to be motivated by that criterion, either primarily or secondarily. Making the distinction between decision-making procedures and right-making properties leaves the objective consequentialist in the same position as the indirect

\footnotesize
\textsuperscript{97} Pettit has stressed in conversation (June 1999) that his version of consequentialism refers to expected outcomes. For arguments in favour of using expected outcomes rather than actual outcomes see Jackson, 1991 and Singer, 1977.

\textsuperscript{98} Railton, 1984, p. 155. This reference to the distinction between truth conditions and acceptance conditions was cited when the motivational test was established in § 1.4.1(ii).

consequentialist. In fact, Railton’s objective consequentialism is just another version of the indirect consequentialism which is in the tradition of the nineteenth century utilitarians.\(^\text{100}\) It falls prey to exactly the same problems as indirect consequentialism that we have already discussed.

(ii) Extreme objective consequentialism

There is another way of understanding objective consequentialism. Extreme objective consequentialism would be the view that consequentialism is a theory about right-making properties and that consequences *never* need to be able to motivate, either directly or indirectly. On this interpretation, consequentialism provides a standard of rightness, and it has nothing to say on the question of how people should be motivated in their decision-making. This is an extreme view because it denies that consequentialism need have any regard for either the psychology or decision-making capabilities of people making actual decisions. As such, it breaks the connection between ethical theory and action.\(^\text{101}\)

Extreme objective consequentialism backs itself into a corner: it ends up being pure in theory, but useless in practice. In Williams’s words, it ‘has to vanish from making any distinctive mark in the world.’\(^\text{102}\) Furthermore, it denies the validity of the motivational test which was established in the first chapter. Extreme objective consequentialism is *too* extreme to be plausible.

2.4.2 Kupperman’s argument

There is one final objection that could be raised. It is this. It might be argued that the motivational test is itself a consequentialist test. Joel Kupperman has an argument which, if correct, would support this objection. He submits the following proposition:

---

\(^{100}\) See n. 5, this chapter.

\(^{101}\) This is the objection that Jackson makes against objective consequentialism. Consequentialism, he says, is ‘a theory about actions, about what to do. ... We have to see consequentialism as containing as a constitutive part prescriptions for action.’ (Jackson, 1991, p. 466.) Jackson’s insistence that it is the task of ethical theory to tell us what to do was referred to when the motivational test was set up in § 1.4.1(ii).

\(^{102}\) Williams, 1973a, p. 135. This quote is cited in Jackson, 1991, p. 314. Jackson thinks Williams’s point is well made against the extreme version of objective consequentialism, but not the moderate version.
Proposition C. In any case in which it is not true that one should do what has
the best consequences, the reason for this has to be based on some judgement of
consequences.\footnote{Kupperman, 1981, p. 305.}

For example, the objection that going about maximising consequences at every
opportunity ‘would be to opt for an undesirable personal character,’ depends
upon considerations of consequences.\footnote{Kupperman, 1981, p. 306.} The idea is that appealing to the kind
of character which one would have if one took consequentialism seriously, is
itself an appeal to consequences. By appealing to consequences in this way, he
claims, the critic of consequentialism shows the truth of consequentialism.

The question is, does Kupperman’s Proposition C apply to the
motivational test? The answer is, it does but only in a trivial way. My
argument runs like this. Consequentialism is limited because it cannot plausibly
motivate. Consequentialist motives are incompatible with many actions which
we would regard as having moral worth. This argument provides an analysis of
what is entailed by consequentialism. It tries to spell out why consequentialism
is an inadequate theory. It appeals to consequences in such a general way that
it is difficult to see how any criticism of consequentialism could not appeal to
consequences. If any alternative to consequentialism or criticism of it turns out
to depend itself on ‘consequences,’ then consequentialism may be true but it is
trivial. For consequentialism to be philosophically significant it must be
possible to formulate an alternative to it.

Conclusion

The question I have set about answering in this chapter is, can
consequentialism plausibly provide motivating reasons for action? I have
argued that it cannot do so in all cases. The problem with consequentialism is
its uncompromising emphasis on agent-neutral values. Some of our most
morally desirable motives arise from agent-relative values. We are often
motivated, not by the net state of affairs which results from our actions, but by
what we must do in the process of bringing about such states of affairs. To
value friendship is not just to bring about a situation in which friendships in
general are maximised; it is to care about the quality of the particular friendship
or friendships that each of us has.
The consequentialist may claim to have a way of incorporating these agent-relative motives. The idea behind indirect consequentialism is that we can keep our agent-relative values but only because they promote agent-neutral values. Agent-relative values, on this account, are only ever justified by agent-neutral values.

The problem with this approach is that some agent-relative values resist being made only instrumentally valuable. For example, if we are only motivated to respect others because doing so serves the higher order motive of bringing about some greater good, then we have misunderstood the nature of such respect. Respect is not always best understood as being an instrumental value. The same is true of friendship. We seek the friendship just for its own sake, not to bring about some greater agent-neutral good. The appeal to indirect motivation does not enable the consequentialist to value respect and friendship in the proper agent-relative way. Even the indirect consequentialist is still motivated ultimately by consequences, if only infrequently. To be sure, the indirect consequentialist is not motivated by consequences in an immediate and clumsy way. But the issue is not whether the motive takes effect immediately or only 'off-line.' The issue is that it takes place at all.

Agent-neutral values have their place within a motivational set. The problem is with the idea that they are the only legitimate values, that the only morally valid standpoint is an agent-neutral one. We shall see in the following chapter that the same problem arises with the agent-relative perspective. Pushed to its limits, agent-relativism breaks down as well. And in the fourth chapter, where we look at an ethic of care, we shall see that, although it can motivate friendship in the proper way, it lacks the impartialist perspective provided by both consequentialism and Kantian ethics. We are edging ever closer to the notion that we require a range of motives.
3

Reasons based on duty

‘Duty, my dear Doctor, before every consideration in the universe!’

Mrs Markleham
David Copperfield

We saw in the previous chapter that the consequentialist looks to the outcome which an action brings about to determine the moral worth of that action. In this chapter I consider another way of determining the moral worth of an action. Kant espoused a way of understanding ethics whereby the reason for the action determines its moral worth. ‘Kantian ethics is moral theory based on intention or volition,’ says Barbara Herman. ‘Things that happen are not themselves morally good or bad, right or wrong: only willings are.’ In particular, for an action to have any moral content it must be performed from a respect for what morality requires. In other words, an action must be motivated by moral duty for it to have any moral worth; it must be done simply because doing it is the right thing to do.

This chapter puts Kantian ethics to the motivational test. This amounts to posing the question, ‘can duty plausibly be a primary motive?’ I argue that, while it may be able to motivate some morally worthy actions, it cannot plausibly be the only motive of morally worthy or admirable actions. There are two problems with the Kantian conception of morality. The first is that it leaves no room for friendship. By this I mean that if someone were to be

---

1 Herman, 1993d, p. 94.
2 The question of whether duty can be understood as a ‘motive’ without misrepresenting Kant’s views on agency is discussed in § 3.1.2.
motivated by duty, as Kant says he must if his actions are to have moral worth, then he would be excluded from having close and worthwhile friendships. This is so because acts of friendship are motivated by affection and concern for the friend, rather than by a rational apprehension of our duty, as required by Kant. The idea that Kantian ethics would ‘squeeze out, ignore, or deform intimate attachments’ is a familiar one in the literature.\(^3\) Kantian ethics has, in turn, been well defended.\(^4\) In this chapter I consider various arguments which may be raised in defence of Kant and argue that they do not succeed. Kantian ethics, with its emphasis on acting for the sake of duty, is quite out of place in the context of friendship.

The second objection takes issue with the agent-relative character of Kantian ethics. We saw in the previous chapter that the agent-relativist only takes responsibility for her own actions, not for the way the world goes. This is especially so in the case of Kantianism. This narrowness of responsibility gives Kantian ethics an unworldliness that is not always desirable. I aim to show how being motivated by a broader sense of responsibility has a place within a motivational set.

In the first section I outline some of the main features of Kantian ethics which are relevant to the ensuing discussion. In the following three sections I argue that Kantian ethics is incompatible with friendship, taking into account various objections. In the fifth and final section I discuss the second reason why Kantian ethics cannot motivate, namely because of its repudiation of the agent-neutral, consequentialist perspective.

### 3.1 Acting from duty, Kantian-style

In trying to articulate and discuss a Kantian conception of ethics, one is immediately faced with the question of the extent to which one needs to be a scholar of Kant’s actual works. Marcia Baron’s and Christine Korsgaard’s approach is helpful in this regard. They understand Kantian ethics to be not simply a scholastic exercise in the history of ideas, but an ongoing intellectual project, conducted in the spirit of Kant’s thought. Such an approach, says

---

\(^3\) Cunningham, 1999, p. 279. For criticisms of Kantian ethics along these lines see Blum, 1980, esp. Chs 1-4 and 1994, Ch. 3; Blustein, 1991, Ch. 18; Cunningham, 1999; Oakley 1990 and 1992, Ch. 3; Stocker, 1987; Williams, 1981a, and Wolf, 1992. For more general criticisms of ethics based on principles see Gilligan, 1982, 1987, 1992; Gilligan and Attanucci, 1988, and Noddings, 1984.

Baron, ‘frees us to develop Kantian views on matters that Kant never addressed, and more boldly, to claim that Kant shouldn’t have said what he did.’\(^5\) One can, however, adopt this approach in varying degrees. While Baron’s writing is peppered by references to Kant’s texts, Lawrence Blum, who sets up a ‘Kantian position’ in opposition to his own views, makes only the vaguest of references to what Kant himself said.\(^6\) The danger with this latter approach is that one may all too easily, through vagueness and generalities, set up a position which nobody actually believes and which just suits one’s purposes (although I am not suggesting here that Blum does make this mistake). In this chapter I adhere fairly closely to Kantian texts, not so much in the name of historical scholarship, but in the name of clarity about a particular view.

Kant’s moral philosophy is very complex and often difficult to follow. A full exegesis is not possible here, nor is it strictly necessary.\(^7\) However, it is important to clarify what is meant by ‘acting from duty’ and to do so is the purpose of this section. In the first part of this section I present a brief exposition of the idea of duty. In the second part I suggest that acting from duty can often provide a plausible motivating reason. The remainder of the chapter is devoted to showing that it cannot always provide such a reason.

3.1.1 Doing one’s duty

Kant says that only an action motivated by duty has moral worth.\(^8\) Why does he hold this view? Convinced that theology cannot provide a plausible framework for ethics, Kant wants to establish an alternative framework which is secure and incontrovertible. Rather than looking outwards to God for ethical certainty, Kant turns inwards, to the nature of reason. He wants to show how reason itself can provide the same kind of certainty in ethics which religion had hitherto provided. As Nancy Sherman says, ‘The historical or “visible church,” as Kant calls it, must gradually yield to the “invisible church,”’ the pure religion of reason that is internal to all of us.\(^9\)

Kant reasons that absolute goodness must underpin ethics. This absolute goodness needs to be independent of the contingencies of time and place. ‘It is impossible to think of anything at all in the world, or indeed even

\(^5\) Baron, 1997, p. 3. See also Herman, 1993f, p. 185.
\(^6\) Blum, 1980, Ch. 1.
\(^7\) For comprehensive interpretations see, Baron, 1995, 1997; Korsgaard, 1996; Nell, 1975; Ross, 1954 and Sherman, 1997.
\(^8\) The following discussion of duty, unless otherwise stated, is gleaned from the *Groundwork for the metaphysics of morals* (G).
beyond it, that could be considered good without limitation except a good will,' Kant famously proclaims at the outset of the *Groundwork for the metaphysics of morals*. Other things, commonly regarded as good, such as virtues of character or gifts of fortune, he goes on to say, are only good if the will of the person who possesses them is good. Courage cannot be good in itself, since a tyrant may show courage. Similarly, wealth is not good without qualification, since wealth may be used for evil purposes. The good will, on the other hand, is good not because of what it achieves or causes, but because of the willing, that is, itself.

Because the good will is the only thing which is unqualifiedly good, Kant says, only the good will is the proper object of moral esteem. Therefore, only those actions which follow from the good will are morally worthy. A good will is one which wills what is right just for the sake of it being right. Thus, the only actions which have moral worth are those which are performed just because they are right. Another way of putting this is to say that only actions performed from duty are morally worthy.

To explain what acting from duty means, Kant contrasts it with acting from inclination. A morally worthy act is not simply one which conforms to duty, but is one which is performed for the sake of duty. There are three cases to be contrasted with acting for the sake of duty:

1. The first is where one acts contrary to duty because the action is useful in some respect. An example (my own, not Kant's) would be where I tell a lie to gain a benefit for myself.

2. The second is where one acts in conformity with duty but one has no inclination to perform the action. As William (W. D.) Ross puts it, 'action A is such as duty prescribes but is done neither because we see it to be our duty, nor because we have an inclination to do it, but because in order to get B, which we desire, we must do it.' I tell the truth, neither because it is my duty, nor because I want to, but purely because telling the truth is necessary to promote my interests.

3. The third is where an action conforms to duty but the subject also has an inclination to perform the action. I tell the truth and I have an inclination to tell the truth, perhaps out of habit, or perhaps out of the satisfaction I derive from doing what duty requires. As with (2), the subject here acts dutifully, but not for the sake of duty.

---

10 G, 393.
11 G, 397.
12 Ross, 1954, p. 15.
In this third case it is difficult to tell what moral worth the action has. We can only be completely sure that an action has moral worth when the inclination is absent, that is, when I tell the truth just because I should, even though I have no desire to do so. Kant gives the example of a philanthropist who, when suffering from grief, loses his usual delight in pleasing others.

Suppose that now, when no longer incited to it by any inclination, he nevertheless tears himself out of this deadly insensibility and does the action without any inclination, simply from duty; then the action first has its genuine moral worth.

It appears as though Kant is saying here that an action has moral worth only in the event that we do not want to do it. That is, it seems as though the absence of inclination is necessary for an action to have moral worth. If this were the case Kantian moral worth would be a very perverse notion indeed. It would mean that not only must we do our duty, but we must find our duty onerous; we must perform it through clenched teeth, so to speak. However, this is not at all what Kant had in mind. In the passage just cited, Kant is comparing two actions by the same person. The first is done from inclination alone: the philanthropist has a sympathetic disposition. The second is done from duty alone. The point is that when the person no longer has the inclination then his action for the first time has moral worth. The action could not have had moral worth earlier, because at that earlier time the action was done solely from inclination. However, the possibility that one could do one’s duty and also take pleasure in one’s duty, has not been ruled out. It is just that when the inclination is absent, the motive of duty is highlighted and made obvious.

Why should this difference in motivation make all the difference for moral worth? To answer this question we need to remind ourselves of Kant’s overall project. He is searching for a secure foundation for ethics. He reasons that inclinations and desires cannot provide such a foundation because it is a contingent matter whether we happen to want this or that. Kant understands the requirements of morality as being categorical; they carry with them an

13 G, 398. We should note here that when one has such ‘deadly indifference’ doing one’s duty may be easy, rather than difficult. This is so because one does not have any desires to struggle against. This point is made by Maria von Herbert in her correspondence with Kant. (See Langton, 1992, pp. 493-494.)

14 For more on this point see Baron, 1995, pp. 148-150 and Herman, 1993a, pp. 18-19.
'objective necessity.'\textsuperscript{15} Therefore, what is required is a motive for moral action which does not depend upon the vagaries of emotion or interest. Duty provides just such a motive. The demands of duty hold fast regardless of our inclinations and, as a reason for action, duty is ever-present. We must do our duty — we must do what is right — and the nature of this ‘must’ is such that it does not depend upon how we feel about the matter, or whether the action is in our interests. If a person acts from inclination, the person is not concerned with whether the action is right or required. She is not acting from respect for the moral law.\textsuperscript{16}

According to Kant, we have an unavoidable interest in our own needs and happiness. We have an \textit{obligatory} interest in doing what morality requires of us, however, whatever our self-interest may be. Kant sees these two demands — self-interest and morality — as being in conflict. Morality acts as a restraint on self-interest. In the same way, the passions contrast with duty as a motive. An action has moral worth only when it is performed from the motive of respect for whatever morality requires in that situation.

We are now able to understand why Kant is not a consequentialist. He believes that only the will of a rational being can be unconditionally good. It follows that only rational beings are capable of producing morally worthy actions.

The moral worth of an action does not lie in the effect expected from it ... For, all these effects (agreeableness of one’s condition, indeed even promotion of others’ happiness) could have been also brought about by other causes, so that there would have been no need, for this, of the will of a rational being, in which, however, the highest and unconditional good alone can be found.\textsuperscript{17}

Because it is not necessary to be a rational being to bring about good consequences, says Kant, the worth of an action cannot be determined by its consequences.

According to Kant, the moral worth of actions cannot depend upon either the contingencies of motive or of consequences. Rather, it must depend upon the willing of a rational being. Now a further question arises: what governs the will? The answer, according to Kant, is not some particular law or principle; it is the moral law itself. ‘Nothing other than the \textit{representation of the}...

\textsuperscript{15} DR, 222.
\textsuperscript{16} For further discussion on why actions performed from inclination can have no moral worth see Herman, 1993a, pp. 2-6.
\textsuperscript{17} G, 401.
law in itself ... insofar as it ... is the determining ground of the will, can constitute the preeminent good which we call moral.\textsuperscript{18} This means that a moral action is an action whose maxim the actor could will to be universal law. This law Kant expresses as an imperative: ‘I ought never to act except in such a way that I could also will that my maxim should become a universal law.’\textsuperscript{19} The moral agent has a duty to act in such a way that she could will all agents to act. This amounts to saying that my own reason for acting (or any agent’s reason for acting) must be capable of being a reason for all rational beings. Moreover, the fact that I could will all agents to so act serves to motivate the action. That is, the moral (categorical) imperative is derivable simply from the idea of moral law, respect for which motivates action. This categorical imperative helps us decide maxims. Maxims are the ‘subjective principles’ which become the reasons for our actions.

It is a characteristic of free agents, says Kant, to be able to choose ends for themselves. This ability to choose is what makes an agent rational and it is what distinguishes him or her from animals and things. This rationality, in turn, is the reason why each person is owed respect, and is the source of moral obligation to others. We respect another person as one who acts on reasons, and who is capable of choosing her own ends. Here rationality contrasts not with irrationality, but with a-rationality. Even if one acts irrationally, one is still a rational agent in the Kantian sense, since the capacity for rationality cannot be forfeited simply because one happens to make bad or irrational decisions.

So far, the requirements of morality have been expressed in only the most general of terms. In \textit{The metaphysics of morals} Kant is more specific. He says that we must adopt two ends: the happiness of others and our own perfection. In promoting the happiness of others we show respect for their autonomy, for their capacity to act upon reasons and to choose their commitments. However, the ends which are chosen must not violate the categorical imperative and must not involve a disrespect for others, so there are limits as to what we can respect in others.\textsuperscript{20}

Kant thinks his ethical theory provides a ‘moral compass.’ It points us in the right direction but, like any compass, it does not tell us exactly what steps to take.\textsuperscript{21} That is the role of judgement. When making a choice in a difficult situation, a person will have already adopted a number of maxims. The

\textsuperscript{18} G, 401.
\textsuperscript{19} G, 402.
\textsuperscript{20} DV, 387-388; 392-393.
\textsuperscript{21} Gregor, 1996, p. ix and G, 404.
decision about how to act will then be about deciding which maxim is the most applicable in that instance, and how exactly to apply it. This is a question of judgement. Consequently, both the *Groundwork* and the later *Metaphysics of morals* focus on principles rather than their application in specific circumstances.

### 3.1.2 Duty as a plausible motive

We are now ready to consider how well Kantian ethics fares in respect of the motivational test. This means answering the question, how plausible is duty as a motive? In the following sections we will look at what motives are ruled out by acting from duty. But first we need to acknowledge the appeal of Kant's theory.

Kant says that he is only describing what 'ordinary reason' and 'natural sound understanding' tell us about morality. How well does his account of moral duty match common understanding?

It is true that we do commonly ascribe a great deal of significance to someone's reasons for acting when we ascribe moral worth to her actions. Suppose a person tells the truth under oath only because she has been threatened with physical harm if she does not. If she had not been threatened, if she had been left to her own inclination, she would have told a falsehood since it was in her interest to do so. Would we wish to bestow moral praise on her for telling the truth? No; she is not telling the truth for the right reason. That much is clear enough. Or, to take another example, suppose an action which benefits another is motivated by an excessive need on the part of the doer to be approved of and accepted. She did not do it because it was right, but to win the good opinion of others. Again, the reason for her action has a bearing on the ascription of the action's moral worth.

The corollary is that we do admire the actions of people who take a stand on principle at great personal cost to themselves. As Kant rightly draws to our attention, when we know that the motive could be nothing other than a sense of what is right, we see most clearly the moral worth of the action. If the person takes the stand simply to attract publicity or just to create a favourable

---

22 G, 394, 397; KpV, 70. The first section of the *Groundwork* concerns the transition from 'common' knowledge of morality to the philosophical. It seems that just as 'The devil can cite Scripture for his purpose,' so too may all manner of moralists cite Common Sense for their purposes. Writing about utilitarianism, Sidgwick also says there is a 'natural transition' from the Morality of Common Sense to Utilitarianism.' Common sense, he says, is 'inchoately and imperfectly utilitarian.' (Sidgwick, 1930, Book 4, Ch. 3, esp. pp. 425, 427.)
impression in the eyes of her betters, the action is no longer so admirable. Indeed, rather than praising the person for the action, we may well view her rather cynically.

To perform an action, not because to do so promotes your interests or makes you happy, not because it fulfils others’ expectations of you, or because it is dictated by unthinking, hypnotic habit, but simply because it is the right thing to do, is truly laudable and noble by any common measure. It is this notion of performing an action just for the sake of the action’s rightness that Kant quite properly takes as a fundamental insight into morality. However, even this idea has its limits, as we shall see.

There is an objection to the motivational test which the Kantian could make right at the outset. It could be said that the motivational test has a bias against Kantian ethics because it makes assumptions about the nature of moral motivation which a Kantian would not accept. We recall that the motivational test claims that external reasons must plausibly be able to be internal reasons. Internal reasons are those which further an element in a person’s motivational set. The problem is that Kant would reject the motivational test since, on his view, moral reasons never depend on the desires that happen to be present in someone’s motivational set. The categorical imperative gives a reason for action which is unconditional and valid for all rational beings. Far from understanding moral reasons in terms of having the right desires, Kant contrasts morality with desire. He wants morality to have force – or command – over our desires.

Herman and Baron make much of this point.23 ‘The problem,’ says Baron, ‘is that the term “motive” suggests causation, as if the motive of duty ... were a force within us that causes us to act accordingly.’24 This empiricist view of agency, she says, ‘has become so standard that many who are sympathetic to Kant, indeed many Kant scholars, use it without noticing that it suggests a very non-Kantian picture.’25

According to Kant, a rational agent always acts on reasons which he has chosen.26 Talking about ‘motives’ suggests that there is something external to the agent determining his action, rather than the agent making a decision to act. Kant’s wariness of the emotions is based on just this idea that, taken too far, the emotions undermine our agency. As Rae Langton says, ‘Inclinations are

---

24 Baron, 1995, p. 189.
26 KpV, 30.
passions in the sense that they just happen to us. And in so far as we let our actions be driven by them we allow ourselves to be puppets, not persons. 27

We need to acknowledge this point. We do not want the failure of Kantian ethics to depend upon the nature of moral motivation; we want it to depend upon considerations of what kind of moral outlook and what kind of life in general the Kantian could have if he acted from duty. Therefore, we need to understand the motivational test in such a way as not to be prejudiced against Kantian ethics. How can we do this if we have already bought Williams’s assumptions about internal reasons being necessarily connected to motivations?

It would help to recall the point of the motivational test. It is to serve as a connector between what the theory prescribes and how it would be if a person were to act as it prescribes. That there should be such a connection should not be something disputed by the Kantian. Kant thinks he is only telling us how ordinary morality – that is, common understanding – is possible. The Kantian accepts the spirit of the motivational test but not the construction of it in Williams’s terminology.

The substance of the charge against Kantian ethics remains the same, whether or not ‘reasons’ and ‘motives’ are understood in terms of desires. The motivational test requires that we ask what would follow if someone were to act from duty in the way prescribed by Kant. For Kantian ethics to pass the motivational test, duty must plausibly be able to be someone’s reason for acting. Whether the mechanism of ‘acting from duty’ involves a desire or not is something we can leave to one side.

By considering what motives would be ruled out in the life of a person who recognised duty in the way prescribed by Kant it will become clear that there is something seriously wrong with the notion of Kantian duty. We will be led to challenge the view that duty is the only morally worthy reason for acting. In the next section we will see that it is only possible to give proper priority to duty (that is, what morality requires) by surrendering our capacity to form and sustain close personal attachments.

3.2 Duty and friendship

What is wrong with acting from duty? Michael Stocker argues, by way of example, that performing actions from the motive of duty, or just because they

27 Langton, 1992, p. 496.
are right, is destructive of friendship. Suppose, he says, you are in hospital convalescing. You are bored and restless until Smith comes along to visit you. You think to yourself what a good friend he is, putting himself out to come and cheer you up.

You are so effusive with your praise and thanks that he protests that he always tries to do what he thinks is his duty, what he thinks will be best. You at first think he is engaging in a polite form of self-deprecation, relieving the moral burden. But the more you two speak, the more clear it becomes that he was telling the literal truth: that it is not essentially because of you that he came to see you, not because you are friends, but because he thought it his duty ... or simply because he knows of no one more in need of cheering up and no one easier to cheer up.28

Stocker says that there is ‘something lacking’ in Smith’s response; he does not have the proper motive. This example targets what Stocker calls ‘the standard view’ that an essential constituent of a morally good act is a morally good intention and that ‘a morally good intention is an intention to do an act for the sake of its goodness or rightness.’29 This ‘standard view,’ with its emphasis on duty for duty’s sake, is distinctly Kantian.

What are we to make of Stocker’s example? In this section I argue that Stocker is right in thinking that the motive of duty is incompatible with friendly motives. In the following two sections I consider how the Kantian might respond.

In the previous chapter we focused upon the non-instrumental nature of friendship. We saw how we value friendship in general, and individual friends in particular, for their own sakes. In this chapter we focus on another aspect of friendship: the importance of affection and the altruistic emotions. To act from friendly motives is to act from the affection and concern we feel for the friend just because she is a friend.

3.2.1 The importance of affection in friendship

We have seen that Kant contrasts duty with inclination and emotion. Performing an action because we are inclined to do so stands in sharp relief to performing an action because it is our duty. Inclination, emotion and desire are

all obstacles to the one moral motive of duty. We have also seen that Kantian
ethics requires a disinterested and impartial outlook. We are only permitted to
act in a way in which everyone could act; we are not permitted to give
ourselves any special priority in our deliberations. Each agent is one among
many and a reason for one must be a reason for all (who are similarly situated).

The reason Kantian ethics sits so uncomfortably with relations of
friendship is because friendship is established and sustained by all the features
which contrast so strongly with what duty is and what it requires of us. Kant
advocates the moral superiority of intellect over desire, of impartiality over
partiality, of universality over particularity. Bonds of friendship are directed
towards particular individuals, to the exclusion of others. We are partial to our
friends. Furthermore, friendships are sustained by mutual feelings of fondness,
affection and fellow feeling. Such feelings then serve as motives.

Let us refer to the person Smith visits as ‘Jones.’ Languishing in his
hospital bed, Jones expresses gratitude to Smith because Jones considers Smith
to be a friend and, as such, to be motivated by feelings of fondness and concern.
When Smith replies that he is merely doing his duty (as a good rational agent)
in visiting Jones, Jones is disappointed and put out because this is not the
motive he either expects or wants from a friend. Moreover, if Smith always had
this attitude to those around him, I suggest he would be incapable of friendship.

It is inappropriate always to act for the sake of duty within the context
of friendship. We want our friends to visit us when we are ill because they
like us and are concerned for us, not because they see that, in visiting us, they are
fulfilling their moral duty, as any rational agent would.

Justin Oakley questions whether even a stranger would want Smith’s
help to be motivated by duty, rather than the altruistic emotions. Oakley
argues that quite generally a person in distress may need action-from-
sympathy, rather than action performed from duty. In suggesting this,
however, Oakley does not give sufficient weight to Kant’s theory of virtue.
Kant does say that we have a duty to cultivate the moral emotions: beneficence, gratitude and sympathy. By cultivating these moral emotions we
transform our ‘animal nature’ in ways that align with morality. These emotions
provide a support for (rather than the foundation of) morality. In reply,
Oakley might say that people may still want sympathy unmediated by duty in
any way. But this is questionable. If one needs a sympathetic response from a
stranger or colleague does it really matter whether her sympathy is

---

31 DV, 452-457; Sherman 1997, Ch. 4, § 4.
spontaneous or deliberately cultivated from duty? It would seem not. But there is not the same kind of indifference in the case of close personal attachments. As a friend of Smith, Jones does not want to be only a recipient of Smith's dutiful benevolence, and if Smith only ever showed such benevolence to Jones, he would not be acting as a friend; he would be acting more like a Good Samaritan.

Kant makes a distinction between two kinds of love, beneficent and pathological love. The Christian injunction to love one's neighbour refers to love of the former kind, a general benevolence towards others. Similarly, when the parent says to the child, ‘You must try to love your brother,’ she is exhorting the youngster to love in the beneficent or non-pathological sense, to show kindness, consideration and the kind of behaviour which is usually associated with pathological love even though the feelings may not be present.

Pathological love, however, says Kant, is more particular in its object and cannot be commanded. ‘Love is a matter of feeling, not of willing, and I cannot love because I will to, still less because I ought to; ... so a duty to love is an absurdity.’ The affection for the friend requires a different kind of sentiment from the generalised benevolence which may be felt for the world at large. The feelings of amity which are found in friendship are what Kant would call pathological; they cannot be commanded by duty. Emotional attachment of the ‘pathological’ kind is necessary for friendship. Such feelings cannot be compelled; they are not obligatory. Quite simply, they are either present or they are not. We either like someone or we do not.

But just because feelings of amity cannot be commanded, it does not follow that such feelings are out of control. It would be a grave mistake to conceive of friendship, or love in general, as consisting in wild, unbridled emotion, always needing to be commanded by the stern voice of reason. Certainly, pathological feelings must be present, but we need to qualify them in two important regards. Firstly, the formation and sustenance of friendships can, to a large degree, be brought under the control of our will. And secondly, friendship places certain demands upon us.

It is commonly said that ‘you can choose your friends but you cannot choose your relatives.’ Usually muttered by disgruntled family members, this adage is generally intended as an observation about the nature of the bloodline. But it also makes a comment about friendship, to wit, that friendship is an object of choice. And this is surely so. We can take deliberate action to

---

32 DV, 401. See also G, 399.
33 For a discussion of the conceptual connection between emotion and friendship see Oakley, 1992, Ch. 2, § 6.
cultivate certain people as friends, to develop friendships, to wind them back or even terminate them. We can do much to bring about the conditions under which 'pathological,' friendly love can blossom or wither.

The second qualification, more germane to our topic, is that feelings of amity need not be the direct motive of each and every act of friendship. Before developing this thought, we need to clarify what is meant by an 'act of friendship.' Broadly speaking, there are two ways of understanding such acts, the second of which shall be our focus. Take the example of Smith visiting Jones in hospital. Smith may act out of a commitment to the friendship itself or he may act out of a direct concern for Jones.

The first possibility has to do with the friendship itself, as something distinct from the good of either of the parties. Suppose Smith and Jones had had a quarrel. To prevent a falling out, Smith may decide to visit Jones in hospital as a gesture of his wanting to sustain the friendship. In such a case, he is acting for the sake of the friendship, to restore their former mutual goodwill and trust. What motivates him is not so much the well-being of Jones, but the well-being of his relationship with Jones. The difference between this kind of act of friendship and the second kind can be understood as the difference between acting for the sake of the friendship and acting for the sake of the friend.34

Let us now consider what it is to act for the sake of the friend. In this instance Smith visits Jones out of direct concern for Jones's welfare. His reasoning would be something like, 'Jones, whom I like and care about, is suffering in hospital. I will visit him to cheer him up.' Smith is, in such a case, caring for his friend just for the other's own sake; as a friend, he wants to do what is best for Jones. Such altruistic caring, says Blum, is definitive of friendship. Blum writes:

For a genuine friend truly cares for the other for his own sake. He is willing to give of himself to promote the other's good. ... He grieves for the friend's sorrows. He is happy for him at his good fortune or successes in valued endeavours; he is sad for him at his losses and disappointments. It is his human growth and happiness which he desires — and for the friend's own sake, not his own.35

34 Stocker makes a similar distinction between acting for the sake of friendship and acting out of friendship. (Stocker, 1981, pp. 754-760.) Similarly, Tom Regan distinguishes between having a duty-to-someone ('someone' being the friend) and a duty-to-friendship-itself. (Regan, 1983, p. 147.)

35 Blum, 1980, p. 75.
A further distinction needs to be made. For there are two ways in which Smith can act for the sake of his friend, Jones. The way Blum describes it in the passage cited above suggests a spontaneity of open-hearted good feelings, an unprompted outpouring of altruistic emotions. The friend 'grieves.' The friend 'is happy' and 'sad,' depending on the fortunes of the other. Such spontaneous willingness on the part of Smith is certainly plausible. However, it would be unrealistic to suppose that all acts of friendship are as unreserved as Blum seems to imply. The second way in which Smith can act for Jones's sake is by reminding himself of his friendship with Jones. In this case, Smith acts for Jones's own sake out of a sense of what his friendship with Jones demands of him.

Why might Smith need to remind himself in this way? He may not have an immediate desire to go rushing off to Jones's bedside. We are often called upon to do things with or for our friends which we would really rather not do. Many commitments are like this. Suppose I have committed myself to developing my musical talents through mastering the trombone. I understand that to realise my full potential I must adopt a daily practice regimen of several hours. Despite my devotion to trombone virtuosity, it is more than likely that there will be occasions when I do not feel like taking my trombone out and running through countless stuporific scales and arpeggios. I would much rather be whiling away my time at the local cafe or loafing about on the beach. Nonetheless, on such occasions, I rally. I remind myself that 'Socrates said practise the art.' If I want the great trombonist within me to blossom and flourish, I have to practise and that is that. Despite wanting to do otherwise, I reach for my trombone case and begin practising. In this scenario it is my commitment to playing the trombone which motivates me, not my immediate feelings about practising.

Acting for the sake of the friend is sometimes similar. Perhaps it was very inconvenient for Smith to visit Jones in hospital that time. After all, he had to postpone a certain rendezvous, for which he had high expectations. Or maybe he was just plain tired. But he came anyway. Why? He felt he should, that it was the right thing to do because Jones was his friend. In this case his reasoning is slightly different from how it would be if he had a spontaneous concern for Jones. It is more along the lines of, 'Jones is a good friend. I'd

---

36 Put like this, it sounds as though Smith is acting in the first way described above, namely, for the sake of the friendship. But this is not so. What motivates Smith here is the care he feels for Jones qua friend. What motivated him in the earlier example was not so much Jones himself, as Smith's friendship with Jones.

37 From the poem by Seamus Heaney, 'A daylight art.'
better go and visit him in hospital. Never mind that I’m feeling tired and it’s inconvenient.’ What does the motivational work here is his commitment to Jones as a friend, which is a deeper motive than how he happens to feel at any particular moment, and whether or not he happens to be in the mood to help Smith.38

The affection which grounds the friendship between Smith and Jones is not a capricious ‘pathological’ feeling. Unfortunately, Kant’s distinctions between duty and inclination and between beneficent and pathological love tends to set up a way of understanding the emotions as unreliable and egoistic. Yet friendship, which is founded on feelings of amity and concern, need be neither.39

3.2.2 The demands of friendship

It is now clear that friendship places demands upon us. As Dr Johnson said to Sir Joshua Reynolds, ‘A man, Sir, should keep his friendship in constant repair.’40 Friendships, like most things to be kept in good repair, require effort and commitment. In visiting Jones because he cares about Jones for his own sake, it might be said that Smith is fulfilling a duty of friendship. However, this duty is not to be understood as a duty he has by virtue of being a rational agent, but a duty he has as a friend.

38 In a similar vein we should also note that the altruistic emotions of which Blum speaks in the context of friendship, are not always immediately forthcoming either. Blum says the friend is ‘happy for him at his good fortune and success.’ In fact, the friend’s immediate response upon hearing of the other’s success may be one of resentment or envy. Similarly, a perverse satisfaction or glee at a friend’s misfortune is not unheard of, at least as an initial reaction. Overcoming such feelings, ‘rising above them,’ is part of the commitment of the friendship too, and is itself a kind of moral achievement. (Iris Murdoch explores this idea of transforming one’s ‘vision’ of another in her example of the mother resolving to see her daughter-in-law, to whom she feels hostile, in a more just and loving light. Murdoch, 1970a, pp. 17-23. For further discussion on this example and the challenge of overcoming one’s darker feelings see § 5.3.2(ii).)

39 Søren Kierkegaard’s distinction between romantic and conjugal love is relevant here. Romantic love occurs when the lover is in the lowest stage of moral development, what Kierkegaard calls the ‘aesthetic stage.’ In this stage the person wants instant gratification of their desires, with little thought of the future or long-term commitment. Conjugal love, however, is what the lover feels when he is in the ‘ethical stage.’ Here the love is more lasting and is grounded in an ethical commitment to the other person. The kind of love that Kant calls ‘pathological’ may parallel romantic love, but it by no means resembles conjugal love, a love which has ethical considerations built into it. (See Kierkegaard, 1946 (first published 1843) and, for an excellent exposition of the difference between the aesthetic and ethical stages, Rudd, 1993.)

40 Boswell, 1968, p. 192. (First published 1791.)
The Kantian maintains that the requirements of morality are discernible through reason, and reason must motivate for an action to have any moral worth. The duties of friendship, however, show that at least some moral demands are discernible by virtue of the relationships in which we stand to others and the bonds of affection. These affections place on us moral demands. This simply means that once one stands in the relation of friendship, there are certain things one ought to do, just by virtue of the relation. The relation carries with it certain responsibilities and, unless these responsibilities are taken seriously, the relationship will wither. The same is true of familial relations, such as the duty of a parent to a child, and other close relationships such as marriage.

We can best make sense of the idea that there are duties of friendship by looking to Blum’s account of friendship. Blum says that it is not enough to understand friendship simply in terms of liking someone. The friend is not just someone with whom we get on and share interests and experiences. Such compatibility and mutuality may be necessary for friendship, but it is not sufficient. The friend is also someone who acts for our good. This idea of the good of another, his or her well-being or interests, is what generates the so-called duties of friendship. Even though Smith did not want to go to the hospital he did it anyway out of consideration for Jones’s well-being. In support of Blum, we may consider the relationships to which friendship stands in contrast. We have already noted a point of difference commonly observed between friends and family members. The difference focused on the element of choice. Friends may also be contrasted with enemies. Indeed, the one may be said to be the opposite of the other. The enemy is your opponent, one who is hostile to you. In contrast, the friend is your well-wisher, one who wants what is best for you. The idea that friendship involves the good of another prevents friendship from being simply a whimsical or egocentric case of pathological love.

The good of the friend can motivate spontaneously, in which case we act directly from the altruistic emotions. Or it may be our recognition of the caring relationship in which we stand to the other person which motivates. The care for the friend may be easy or it may require work, in the sense of overcoming contrary inclinations. These are the two ways in which one can act for the sake of the friend. When one is motivated in the second way, by a

41 Blum, 1980, Chs 3 and 4.
42 To be sure, the altruistic emotions are not infallible motives of right action but, as Oakley argues, neither is duty infallible. One may act from duty but have false beliefs, make calculative errors or have poor judgement. (Oakley, 1990, pp. 445-452.)
commitment to the other person, one may be said to be acting from the duties of friendship.

It now appears as though we have identified two kinds of duties, those performed as a Kantian rational agent and those performed as a friend. The point is that duties of friendship are not just instances of Kantian moral duties. The use of the term 'duty' here is confusing. There are two differences between the duties of friendship and Kantian moral duties: (1) how these duties are arrived at and (2) what motivates us to perform them. In respect of (1), Kantian duties are arrived at by considering the moral law, that is to say, how we could will all rational agents to act. The duties of friendship, on the other hand, arise from the commitment we have to the friend’s good, a commitment we have because we care about the friend. Secondly, we are motivated to perform Kantian duties by the fact that they are duties. Kantian duties are categorical; we are obliged to fulfil them. Friendship duties only arise in the context of our particular and contingent attachments to individuals. The friendship duties are motivated by the fact that the duty pertains to the friend. They are performed for the sake of the friend, rather than for the sake of the duty.

The Kantian would understand Smith’s duty to visit Jones in terms of the maxim, ‘visit one’s friends when they are ill.’ Smith has adopted this maxim which, of course, he can will all agents to adopt. When Jones is hospitalised Smith judges that he (Smith) should act on this maxim. The Kantian may argue that the duties of friendship are therefore no different in principle from other duties. To see what is wrong with this, we have to ask ourselves why one would adopt the maxim, ‘visit one’s friends when they are ill,’ in the first place, if indeed one adopts it at all. One adopts it out of consideration for one’s friends. It is an expression of the friendship, and a recognition of the demands of care which the relationship places upon one. We return here to the point about moral requirements arising from feelings, in this case feelings of amity. One does not adopt the maxim because one respects the other person as a rational agent. One does, of course, respect the friend, but one loves her as well. One respects all rational agents, but loves only a few friends. The duties of friendship arise from the love, not the respect.

Let us suppose Smith had succumbed to his fatigue that night and let us further suppose he felt guilty the next day. What would be the source of his guilt? That he had not acted on a universalizable maxim? That he had not shown due respect for the autonomy of a fellow rational creature? Hardly. He would feel, rather, that he had let Jones down, that he had not risen to the demands that Jones’s affections had placed on him. His guilt cannot be
properly understood without referring to his friendship with Jones, and the fact that he likes Jones.

The Kantian might reply that these feelings of guilt are neither here nor there, that we cannot place too much store by these feelings since one can feel guilt inappropriately and on the wrong occasions. Indeed, one can have misplaced guilt, but we are discussing here what it is intelligible to feel guilty about, not whether on any particular occasion the guilt is justified. And what it is intelligible to feel guilty about reveals something about the nature of the wrong committed.¹⁴³

Let us summarise what has been established. The affections we have for others carry with them certain demands. The demands arise because friendship, as we saw, has to do with wanting what is best for someone for whom we feel fondly. Actions for the good of the friend are not performed for the sake of duty but for the sake of the person who is the friend. In acting for the sake of duty we are motivated by our respect for what morality requires, as indicated by the categorical imperative. What motivates us in acting for the sake of a friend is our concern and care for the other person. Without this motive the friendship is not possible.

The point of this section has been to show that the Kantian understanding of duty could not plausibly motivate acts of friendship, since affection and emotion are ultimately the appropriate motives for acts of friendship. I shall now consider two responses which might be made on the part of the Kantian.

3.3 Duty and affection as coexistent

There are two ways in which a Kantian could reply to the charge that duty is incompatible with friendship. The first of these is to argue that it is possible to act from both duty and friendship at the same time. The second reply is to agree that friendship is excluded by duty but to deny that such exclusion is a problem. After all, goes this second reply, what is being excluded is not friendship itself, but the moral worth of friendship. These two replies are the subject of the following two sections.

Let us now consider the first of these replies. To defeat the objection that acting from the motive of duty is incompatible with friendship, the Kantian requires an account of action which has the following two features.

¹⁴³ See Gaita, 1991, Ch. 4 for a discussion on the moral significance of remorse.
Firstly, the action must be performed from friendly motives and, secondly, the action must have moral worth. How could such an account be possible within a Kantian framework? Herman claims to have provided just such an account.

Herman’s idea is that when love and duty are both present and both capable of being primary motives and the recipient would prefer to have the action done from love, then it is perfectly legitimate for the agent to defer to his feelings and act from love. Her point here is that the agent’s good will is not diminished by his acting from love.

Suppose C’s friend Y needs help. C has a settled and sure commitment to helping others, and his personal feelings for Y also move him to help. There is no moral reason why C would have to help Y beneficently: that is, help him with a sense of doing what any moral agent is required to do for any person with such need. He would be acting no less well in deferring to his feelings ... As C was ready to help without regard to his feelings, and so from the motive of duty, he fully satisfied the moral requirement of beneficence. We might say: given his feeling for Y, C had no need to act out of a sense of duty.44

Herman says that it is not morally required that C act from duty. It is permissible that he, C, act from love because C has already proven that he has a good will. The assumption is that it is sufficient to have a good will; one does not need to act upon it. But ‘sufficient’ for what purpose? Sufficient for the action to have moral worth or sufficient simply for it to be permissible? It must be the latter, since an action can have moral worth only in those cases where duty is the primary or initiating motive. In cases where we ‘defer’ the motive of duty, duty must play a passive secondary role as a motive. Duty would, of course, be the primary motive were the love absent. Herman says she has shown that ‘it is not morally required that we always set the motive of duty between our feelings and our responses to others.’45 Friendship and duty can co-exist as motives so long as duty is a secondary, regulative motive.46 But actions performed when duty is a secondary motive do not have moral worth.

But suppose we were generous about Herman’s idea that we need not be obligated to act from the motive of duty. Suppose we were to grant that the mere presence of a good will is sufficient not only for an action to be permissible but also for the action to have moral worth. That is to say, imagine the following scenario. C is motivated by both the duty of beneficence and by

---

44 Herman, 1993b, p. 36.
45 Herman, 1993b, pp. 36-37.
46 Baron (1997, Ch. 4) has a similar argument.
love. Either of these motives is strong enough to be a primary motive. However, C ‘defers’ to the motive of love. By being generous to this account, we say that C’s action has moral worth. In saying this, we may be departing somewhat from the spirit and the letter of Kant’s own writings, but let us disregard that point for the moment. Here we maintain that if the motive of duty could be a primary motive, that is, it is strong enough to motivate by itself, then the action is morally worthy. Have we not then found a way of satisfying the two conditions which we said earlier need to be met if the Kantian position is to be defended? We recall that the two conditions were (i) that the action be performed from friendly motives and (ii) that the action have moral worth.

Such serpentine manoeuvring can indeed deliver us a way of reconciling morally worthy actions with friendly actions. However, the Kantian has not yet won the day. To understand why this revision of Herman’s argument cannot fully rescue Kantian ethics from the objection at hand, it would be helpful to consider all the possible ways in which the motives of duty and affection could be present in friendship. For the sake of simplicity, I refer here to all friendly motives as ‘love.’ There are five ways in which duty and love could feature in the context of friendship.

(1) **Duty only is present. Duty motivates.** In such a case as this, one performs one’s duty for its own sake; the fact that the action is directed towards a friend has no motivational significance. Such an example might be telling the truth to a friend. One tells the truth because it is one’s duty to do so and one would do it regardless of whether the recipient were a friend or not. In this case the action has moral worth.

(2) **Duty and love are both present. Duty motivates.** This second possibility depends upon it being possible to have an inclination without that inclination motivating. When two motives are present like this the action is ‘overdetermined.’ In this case the love is present but it does not do the motivational work. Because the action is performed from duty, it has moral worth. Baron gives a nice example of this second possibility.

Kevin agrees to pay Karen a set sum for painting his living room. She paints it as planned and does an excellent job, and he pays her the sum

---

47 We will also leave aside the puzzle about the mechanism of ‘deferring’ motives. (It is difficult to see what this might mean other than that the motive deferred to is the stronger motive.)
agreed on. Why does he pay her? Because he agreed to ... He also is happy to do so, since she did such a good job. But his “inclination” to pay her does not contribute to his action of paying her. He pays her the set sum not because he wants to, or because he wants to and because she did such a good job, but because that was the agreement.48

Many obligations to friends fall into this category.

(3) Duty and love are both present. Both motivate. There are two possibilities.

(a) Both motives are sufficient on their own to motivate action, but because the action relates to a friend the agent ‘defers’ the duty motive. Let us grant that actions of this kind have moral worth. This is the claim made by the revised version of Herman’s argument.

(b) Neither motive in itself is sufficient to motivate action, but together they are. Because the love is necessary for the action to be performed the action cannot be said to have moral worth.

(4) Duty and love are both present. Love motivates. Because in this case the action is performed from love, the action has no moral worth.

There are some cases where if the love is present then it must do the motivational work. That is to say, it is not enough that inclinations be present; they must also motivate. Consider the character St. John from the novel, Jane Eyre. St. John is motivated by duty. To be sure, it is Christian duty rather than Kantian-style moral duty which motivates him, but the distinction is not important for the point at hand. He desires only to do good in the world through carrying out his duty. To that end, he proposes marriage to Jane. He makes no pretence of having any feelings of love towards her; he thinks only that a marriage between them will enable him to carry out his missionary duties better in India. He tells Jane, ‘You shall be mine: I claim you – not for my pleasure, but for my Sovereign’s service.’49 It is his very absence of feeling, and attachment to duty, which makes him such an unattractive character and which renders it impossible for Jane to accept his proposal.

48 Baron, 1995, p. 152.
49 Brontë, 1996, p. 448. (First published 1847.)
We can imagine a character similar to St. John, but who also has a warm and affectionate character. He is motivated by duty in his proposal to Jane, but he also loves her. Would we see him as so objectionable? There would be something very odd about such a person for him not to be motivated by his affection. Imagine him declaring to Jane, ‘even though I love you, I propose to you because I see it as my duty to have a wife such as yourself.’ This would be a very strange speech and scarcely one which would induce Jane (or any other normal woman) to accept the proposal. Not only should affection be present in such circumstances, it should motivate. If the love does not motivate, one wonders whether the love is really present at all. The example of Smith and Jones may also fall into this category. If Smith visits his friend, Jones, from a sense of beneficence, that is, he just does what anyone in his situation would do, then he is not visiting Jones as a friend.

Only love is present. Love motivates. In this case the action is performed purely out of love. Duty is absent; one performs an action only because the recipient is a friend. The case of Smith and Jones could be just such an example. Smith is not obliged to visit the sick in hospital (particularly since the hospital is so difficult to get to); he only visits Jones because Jones is a friend. If Jones were not there the idea of going to the hospital would never have entered his head.

These, then, are the five ways in which love and duty may interact in the context of friendship. We are now able to see better why even the revised version of Herman’s account is so limited. As we proceed down the list from (1) to (5) the motivational capacity of duty becomes less clear and less significant. At the same time, the weight given to the motive of love increases. Even if Herman’s revised option – (3)a – is plausible, it does not exclude the fourth and fifth possibilities and these last two are explicitly acts of friendship. All Herman has shown is that it is sometimes possible to be both a good Kantian agent and a friend. But she has by no means covered all acts of friendship, and she has not covered the acts which really count. The difficult cases are those where one acts solely from friendship, where one acts in a certain way towards another just because that person is a friend. To the extent that one behaves just as anyone would or should in that situation, one is not behaving as a friend, one is behaving as a Kantian rational agent. We can grant that there may be some area of overlap, but we should remember that friendship is founded and sustained by the affections. If there were no actions
at all performed from the motive of love, there could not be a friendship in the first place. It is actions of the fourth and fifth kind that enable us to regard others as friends and to be friends ourselves.

We are left to conclude that the motives necessary for friendship are quite different from the motive required for moral worth in the Kantian scheme. Friendship and Kantian duty cannot be reconciled.

3.4 Friendship as outside the scope of morality

At the outset of the previous section I suggested there were two ways in which the Kantian could reply to the idea that the Kantian moral agent would be unable to have friends. The first reply was to argue that duty and friendship could co-exist as moral motives. Let us now consider the second reply. The Kantian could say that, while friendships are convivial enough, even conducive to personal happiness, their formation and sustenance is not morally praiseworthy. Indeed, that portion of it which consists in a ‘pathological’ liking is, by virtue of its very pathology, non-moral. Friendship simply falls outside the scope of morality.

When the Kantian claims that friendship falls outside the scope of morality he means that it does so in the same way that ‘such qualities of temperament as courage, determination and constancy of purpose’ do. That is, like power, riches and wit, friendship is ‘undoubtedly good and desirable’ in many respects, but because it is not unconditionally good, it is not the proper object of morality.\(^50\)

This idea of friendship falling outside the scope of morality might be made plausible if we once again draw on the distinction between primary and secondary motivation. Kant says that the motive of duty is ubiquitous. However, the ubiquity of morality does not require a maximising approach to duty. We are not required to seek out burning houses and other dangerous situations just so that we can show how morally worthy our actions are when we overcome our fear and act as duty commands.\(^51\) If we employ Herman’s distinction between primary motives and secondary or limiting motives, we can see how duty may be present without it being necessitated that every action be

---

\(^{50}\) G, 393.

\(^{51}\) A similar point could be made in defence of Kant against Susan Wolf’s criticism that a Kantian saint would be a very unappealing person. (See Wolf, 1982.) In criticising the Kantian-style saint as having an unenviable personality, Wolf assumes Kant has a maximising approach to doing one’s duty.
performed from the motive of duty. ‘As a limiting condition, the motive of duty can be present in (or satisfied by) an action, and yet that action may have no moral import.’ Thus, so the argument goes, in friendship we are able to act from affection provided that we do not transgress the requirements of duty. Acts of friendship, while not having any moral worth, would nonetheless be permissible.

There are two ways to respond. The first is to challenge directly the legitimacy of denying that friendly motives are moral motives. The charge against Kantian ethics, according to this line of reasoning, is that Kantian ethics fails the motivational test because it rules out friendly motives as moral motives. That is to say, it prescribes too narrow a range of moral motivations.

The second way of responding is to leave it an open question as to whether friendship falls outside the scope of morality. This way we do not challenge the moral status of friendship; we challenge the kind of friendship which it is permissible for the Kantian to have. The idea is that the friendship which would be open to someone who was genuinely motivated by duty, with all that that entails, would only be capable of a very impoverished kind of friendship indeed.

3.4.1 Moral features of friendship

If a moral theory places friendship outside the scope of morally worthy actions, what should we conclude? Should we conclude that, as the theory says, friendship has no moral content, or should we conclude there is something wrong with the theory? Kant, as we have seen, thinks the former; friendship, like all desires, is subject to the requirements of morality. Friendship needs to be carefully regulated by the principle of respect. Our feelings for each other are wont to lead us either to disrespect the other or to be shown disrespect from the other. We must be on our guard against friendship, since, like all inclinations, it can lead us morally astray.

There is another view, however, beginning with Aristotle, which understands friendship as falling squarely within the scope of morality. Aristotle’s idea was that one cannot live up to the ideals of friendship without displaying moral virtue. Indeed, one’s capacity for friendship is only as good as one’s virtue.

---

52 Herman, 1993a, p. 17.
53 Herman (1993f) argues along these lines.
54 DV, 470-471.
55 NE, 1155a-1172a.
In more recent times the idea that one displays moral qualities in friendship has regained favour.56 There are two related ways in which friendship may be understood as having moral currency.57 The first is that it involves a notion of the other person's good. We saw earlier that part of what prevented friendly feelings from being capricious was that the feelings are directed toward the good of the other person. The second is that it is based upon the altruistic emotions.58 Let us now develop this idea of acts of friendship having moral worth by way of examples.

Dean Cocking and Jeanette Kennett’s recent paper, ‘Friendship and moral danger,’ shows how friendship and (Kantian) morality might be contrasted with each other.59 Cocking and Kennett, like Kant, understand friendship as being a non-moral good. They make their point by way of an example from the Australian film, *Death in Brunswick.*60 Carl, a ‘weak, vain, and disorganised’ character, through a string of unfortunate circumstances, calls upon his friend Dave, ‘an easy-going family man,’ to help him move a dead body in the middle of the night. Carl has accidentally killed the man with a kitchen fork in self defence. If the body is not moved, Carl is certain he will go to jail. Dave reluctantly agrees to help him. This story shows, say Cocking and Kennett, how the demands of morality can conflict with the demands of friendship.

In contrasting morality with friendship, Cocking and Kennett assume a Kantian framework of morality. Morality tells us that there are certain principles which should regulate our conduct. (Illegally moving dead bodies, then lying to the deceased’s family about it, would be something prohibited by morality.) Friendship can pose a threat to our moral resolve, since friends can exert pressure on us to do the wrong thing. Kant, himself, is wary of friendship for precisely this reason.61 Both Kant, on the one hand, and Cocking and Kennett, on the other, can agree that ‘the good of friendship does not sit well within the moral framework.’62

---

57 This was strongly suggested in § 3.2.
58 Blum, 1980.
59 Cocking and Kennett, 2000. Frankfurt and Wolf would also be sympathetic to this contrast, juxtaposing as they do the demands of love with those of morality. (Frankfurt, 1988, p. 268, 1999b, 1999c; Wolf, 1992, pp. 254-256.)
60 Ruane, 1990.
61 DV, 469-471.
62 Cocking and Kennett, 2000, p. 296.
But why should the demands of friendship be understood as having no moral content? This is to assume the very point at issue, namely the scope of morality. Much depends upon how situations such as that involving Carl and Dave are portrayed. The first point to be made in discussing this example is that *Death in Brunswick* is a comedy, a spoof on the classic film, *Death in Venice*. The scenes in question are meant to be lighthearted and humorous, rather than evocative of any moral issues at stake. Nonetheless, Cocking and Kennett, in their explanation of the film, do not do justice to the way in which Dave responds to Carl’s request. When Dave arrives at the kitchen where Carl has accidentally killed his victim, Dave is most reluctant to become involved. He urges Carl to go to the police. Dave tells Carl that he cannot get involved because he, Dave, is a family man. Carl is panicky and desperate. He pleads with Dave to help him dispose of the body. His ultimate appeal is to their friendship. “Dave, you’re my friend.” This entreaty moves Dave. His friendship with Carl places certain demands upon him which he cannot ignore.

To say that the friendship places moral demands on Dave is not to commit oneself to the view that Dave, all things considered, did the right thing in helping Carl. For it may well be that, on this occasion, other moral considerations outweighed the demands of friendship. What is at issue is the way in which we should understand Dave’s choice. Dave responded as he did because he cared about what would happen to Carl, he understood Carl’s vulnerability and innocence. He wanted what was best for Carl. He was motivated by the altruistic emotions. Furthermore, had he not helped Carl, he would very likely have felt guilty, that he had let Carl down. All this suggests that the choice that Dave faced was between competing moral claims.

Consider another example, this time from the 1996 film, *Sleepers*. A priest, Father Bobby, finds himself in a difficult situation. He is asked to perjure himself by providing a false alibi for two men charged with murder. These two men have vengefully killed a man. The dead man had abused them emotionally, physically and sexually when they were in a boys’ remand home in their youth. The men are well known to Father Bobby, since he had been a father figure and mentor to them when they were growing up on the streets of his tough working class New York parish. He had been a stalwart support while they were held in remand. Father Bobby is himself a reformed criminal. What is more, he had also been sent to the same remand centre as a youth so he fully understood the horror of what the boys had been forced to endure. What should he do? Should he tell a lie under oath so the men may be set free? Or

---

63 Levinson, 1996.
should he take no part in the scheme and see them punished for a crime they had been provoked to perform?

The obvious way of describing this situation would be to say that the priest has two different kinds of moral demands placed upon him. On the one hand, he has a duty to tell the truth under oath. This duty might be understood as a requirement of Kantian morality. On the other hand, he has a duty to help those he cares about fight their injustices. This latter duty could be understood as a duty of friendship. Although, strictly speaking, the relationship between Father Bobby and the two accused men is not so much a friendship, as a paternal interest, nonetheless the demands placed on Father Bobby by his relationship with the men are similar to those generated by friendship. He feels a particular obligation to them because he knows them personally, is fond of them and cares for them.

Father Bobby, aware of the seriousness of the choice he needs to make, takes time out alone to think it over. That he agonises over this decision, and that it is quite understandable to us why he should be agonising, suggests that the choice he needs to make is between two kinds of moral requirements, one grounded in the demands of impartial reason, the other in the demands of contingent relationships of affection. We know that whichever way he chooses, there will be a loss, that he will have failed, through no fault of his own, to fulfil an obligation.

Cocking and Kennett’s way of understanding such a situation in terms of morality versus non-morality has less appeal, I suggest, in this example. To describe Father Bobby’s sense of commitment to the welfare of these men as having no moral dimension would be to understate, even trivialise, the care he felt for them. Similarly, the idea that Smith’s actions have no moral worth when he puts himself out to visit Jones is to diminish the value of what Smith does.

Because Kant understands morality as grounded in human agency and respect for agency, important dimensions of our moral relations with others are not given due recognition. In particular, actions done from love, other than beneficent love, are disregarded. In the next chapter a moral theory will be discussed, namely an ethic of care, which does present a moralised account of friendship and other close relationships. I hope to have indicated that a good case may be made for understanding friendship as a moral good, as opposed to a non-moral good. However, the argument against Kantian ethics does not depend upon us understanding friendship in this way.
3.4.2 The calibre of Kantian friendship

Suppose we grant the idea that friendship is to be contrasted with morality, rather than being on a moral par with it. The question now is, what kind of friendship would a Kantian be capable of? The answer is that a true Kantian would only have a very limited capacity for friendship, if there were any capacity at all. There are two related reasons. The first is that the Kantian friend would have an emotional life that was too impoverished. The second reason is that the Kantian friend would remain too detached.\(^{64}\)

(i) The emotional life of the Kantian friend

It might be said in defence of Kant that he thinks highly of friendship.\(^ {65}\) After all, he describes friendship as a "union of two persons through equal mutual love and respect." Because friendship requires that we have a good disposition towards each other, "striving for friendship ... is a duty set by reason, and no ordinary duty but an honourable one."\(^ {66}\)

The fact that Kant can say we have a 'duty' of friendship is worrisome. For we know that pathological love cannot be required. It suggests that Kant has an odd conception of friendship – one that denies any pathological dimension. But there is no escaping the pathological aspect of friendly love, and it is the very pathology of it that makes it so pleasing and sought after. In friendship, weary from the necessities of duty, we enter a zone where we can let our moral hair down, so to speak, and indulge our pathological preferences.

---

\(^{64}\) For the sake of completeness, it is worth mentioning other ways of arguing that a Kantian would make a poor friend. Cocking and Kennett say that our friends are 'inherently likely to lead us into moral danger.' (Cocking and Kennett, 2000, p. 296.) This was the point of their story about Dave and Carl. Kant would say that where the demands of morality conflict with the demands of friendship, morality must always win out since morality places a *categorical* demand upon us. But would we wish to say that every single instance of placing the demands of friendship over the demands of an impartial morality was wrong? This would probably rule out most ordinary cases of friendship, where displays of partiality and minor rule-bending are common occurrences. Anthony Cunningham also argues that one cannot be a good Kantian and have intimate attachments at the same time. His argument is a variant of Wolf's 'Moral saints' argument: Kant's imperfect duty of alleviating suffering is so demanding that it leaves no scope for intimate attachments. (Cunningham, 1999, p. 290.)

\(^{65}\) Allen Wood, for one, has this view. He says that we would need to go back as far as Aristotle to find a major philosopher for whom friendship is so important. (Wood, 1999, p. 275.) Kant discusses friendship in DV, 469-473; LE, 422-430, 675-686, 696-697.

\(^{66}\) DV, 469.
for those individuals we *like*.\(^{67}\) There can be no moral obligation to have friends or seek them.\(^{68}\)

Kant says we have a duty of friendship because the kind of love he has in mind regarding friendship is moral love, beneficence. Indeed, this is the point Korsgaard makes when arguing that there is room in a Kantian life for personal relations. Her understanding of personal relations is that they are just like relations with any other rational agent, only more intense, more focused. She says

> Anyone must tell the truth when the circumstances call for it, but between friends there is a presumption of intimacy, frankness, and confidence. Anyone must help another in need or emergency, but friends promote each other’s projects as routinely as they do their own. ... To become friends is to create a neighborhood where the Kingdom of Ends is real.\(^{69}\)

A friend, in other words, is a rational agent, only more so. What is right about this passage is that there is a presumption of intimacy on the part of friends. What is wrong about it is the assumption that friendly love is identical to benevolent love in kind and that the only difference between the two is one of *degree*.\(^{70}\) It overlooks the point that friendship is not just a mutual benevolence society founded on a moral duty to help others. Friendship is based on affection and the attractions of what Kant calls a ‘pathological’ liking for someone.

Kant’s ‘duty friendship’ is emotionally bereft. His portrait of the emotional life of the friend is revealed in the following passage. This passage is not unexpected, given Kant’s views both on the emotions and the nature of friendship. Influenced by the Stoics, Kant writes,

> It was a sublime way of thinking that the Stoic ascribed to his wise man when he had him say “I wish for a friend, not that he might help *me* in poverty, sickness, imprisonment, etc., but rather that I might stand by *him* and rescue a

---

\(^{67}\) Not only are the necessities of duty wearying, but so too are the expediencies of self-interest. This is the other refreshing aspect of the friendship zone: we do not seek any worldly advancement from the friend; only the friendship itself. See § 2.3.1.

\(^{68}\) None of this is to say that friendship does not place constraints on us of a moral kind. But the constraints arise out of the *friendship*, not out of the other person’s autonomy.


\(^{70}\) Korsgaard admits as much when she says that the degree of reciprocity is the difference between moral and personal relationships. Korsgaard, 1992, p. 309.
human being." But the same wise man, when he could not rescue his friend, said to himself ‘what is it to me?’

The last sentence is the most objectionable part of this passage: if one cannot rescue the friend, one should simply turn away. Following the passage cited above, Kant tells us that when I allow myself to be ‘infected’ by the suffering of another, even when I cannot help him, I suffer unnecessarily. ‘Then two of us suffer, though the trouble really (in nature) affects only one.’ The idea is that there is to be no excess of emotion. Emotions are needed as a support for doing one’s duty, but that is all.

Even Baron, who defends Kant against the charge of emotional coldness, finds this excerpt troubling. This ‘dreadful’ passage, she tells us, accounts for the ‘almost’ in the title of her book, Kantian ethics almost without apology. She tries to soften the excerpt by suggesting, along with Susan Mendus, that Kant may mean to say here that if one is in a position where one cannot help the friend, then one should turn away. One should be practical and not give way to ‘romantic notions that the more we weep and the more overcome by sadness we are ... the better our characters are.’ Sympathetic feeling only has value when it can be used to help someone in a practical way. Even if Kant’s target here is some romantic notion of suffering, he overshoots his target in this extract and in those similar to it.

Kant means to go further than simply saying that if one cannot be practical, one should turn away. His point is that one should turn away emotionally. One should ‘shut down.’ But is this really an admirable ideal? It is useful to consider Williams’s essay on regret in this context. In cases where someone believes there is an obligation to take two conflicting courses of action, he may feel regret at not taking one or the other of them. Even though he is convinced he made the right choice he feels regret ‘for what was missed.’ Decent human beings, says Williams, are disposed to have exactly this sort of

71 DV, 457. Kant’s ethics was ‘in essence a return to Stoicism.’ (Ross, 1954, p. 92.) Certainly, Kant was impressed by the unworldliness of the Stoics, their imperviousness to the contingencies of fortune and their defiance of moral luck.


73 Baron, 1995. The reference to ‘dreadful’ is p. 221; the reference to the title is p. 9.


75 Baron, 1995, p. 209.

76 Baron 1995, p. 9. Sherman agrees that Kant is repudiating sentimentalism in this passage. (Sherman, 1997, pp. 151-153.)

77 Williams, 1973b, p. 170. Williams’s observations about regret were referred to in the Introduction.
reaction. Richard Hare goes further and says that we would think the worse of someone who did not have this reaction. For example, he says, suppose I have been brought up with the belief that one ought not to break promises. Now suppose I find myself in a situation where the circumstances require me to break a promise. 'I cannot then just abolish my past good upbringing and its effects; nor should I wish to.'

These observations seem right. The moral emotions such as regret, remorse and compunction bear witness to our sense of right and wrong. When we find ourselves in situations where we are unable to do what we think we should, the regret we feel is evidence of our understanding that there has been a loss. Along similar lines, Jeffrie Murphy argues that, at least in some circumstances, resentment at wrongdoing 'against oneself is a morally desirable emotion. It shows that one has a sense of self-worth. If one had no resentment when one had unjustly suffered at the hands of another, it would be questionable whether one had any self-respect.

The emotions, then, can reveal much about what we value. The same is true for friendship. Just as we feel regret about a certain decision, even as we know we made the right decision, so too does an admirable person feel regret that she could not help her friend, even though she knew that there was nothing she could do to help. She wanted to help. Why? Because she cares about the welfare of her friend. The regret that she feels upon the realisation that she is unable to help is evidence that she does, in fact, care. Regret is not the only emotion which reveals an attachment. Sorrow, grief, despondency would all fall into the same category.

To experience these emotions does not make one an hysterical romantic. Such emotions simply reveal an attachment to another person. The question now is, could one have the relevant kind of attachment without the corresponding emotions of regret, sorrow, grief and so on? Could one love the friend, yet not feel grief upon her death? It is doubtful whether it is possible to pull these two emotions apart. As Hare says, 'feeling guilty is inseparable from ... thinking that I ought not.' In the same way, feeling regret about being unable to help is inseparable from feeling that one wants to help. One wants to help, in this case, because the other person is a friend.

---

78 Williams, 1973b, p. 173.
79 Hare, 1987, p. 209.
80 Hare, 1987, p. 209.
81 Murphy, 1988, pp. 16-19.
82 So, too, would positive emotions such as elation, satisfaction and happiness. However, these are not Kant's targets.
According to Kant, there is to be no emotional excess. One must remain self-contained and self-governing. However, this emotional self-containment stifles the impulse for friendship, for emotional attachment to others. Curtail the emotions which signify attachment to others – and the possibility of friendship is itself destroyed.

(ii) The detached stance of the Kantian friend

This brings us to the second reason as to why the friendship which is permitted under the Kantian regime is problematic. The Kantian agent remains too detached from others to form proper friendships.

When writing about friendship, Kant’s tone is one of wariness. The problem is that friendship is founded on emotional response, and ‘emotion is blind in its choice, and after a while it goes up in smoke.’ One may have friendships, but only with caution. He tells us that it is not wise for friends to be too familiar with each other. ‘The principle of respect requires them to stay at a proper distance from each other.’ Friendship involves a constant tension between love and respect, Kant says. Love draws the friends together, while respect puts them at a distance from each other. ‘Love can be regarded as attraction,’ says Kant, and ‘respect as repulsion.’ The principle of respect limits the intimacy which friends should have.

In taking issue with Kant about the need for detachment, we should first acknowledge what is right about his view. Common sense tells us that one needs to exercise judgement with others when it comes to developing intimacy with them. If one friend becomes more attached because of an emotional neediness, there is an imbalance and, possibly, a loss of respect for the one who is more attached. Keeping a distance from the other person is also a precaution against their smothering you, taking you over and imposing their wishes upon you. A distance between friends not only protects the friends from emotional vulnerability, but ensures that there is no imposition on one by the other.

We can admit the need to exercise this kind of judgement, however, without denying ourselves the possibility of finding friends where there is not such a need for distance. The degree of distance will vary, according to the circumstances and capacities of the parties involved. We need an instance of

---

84 DV, 471.
85 DV, 470. Similar thoughts are expressed in LE, 682.
86 DV, 470.
friendship which breaks the rules, as Kant prescribes them. Michel de Montaigne provides just such an example in his essay on friendship. 87

With eloquence and poignancy Montaigne writes of his friendship with Étienne de la Boétie. The men met in their twenties but, owing to la Boétie’s premature death, their friendship was cut short after six years. This was no ordinary friendship. It was ‘complete and perfect.’ What made it complete and perfect was the absence of the distance and respect of which Kant speaks. There was a ‘seamless union’ in their friendship. Montaigne tells us that at their first meeting, ‘we found ourselves so captivated, so familiar, so bound to one another, that from that time nothing was closer to either than each was to the other.’ 88 For Montaigne, there was no question of either of them leading the other morally astray. Montaigne felt as certain of la Boétie’s will as he did of his own. Knowing that his friend was ‘guided by virtue and governed by reason,’ Montaigne felt an absolute trust and confidence in his friend. 89 This is how he puts it.

Our souls travelled so unitedly together, they felt so strong an affection for one another, and with this same affection saw into the very depths of each other’s hearts, that not only did I know his as well as my own, but I should certainly have trusted myself more freely to him than to myself. 90

The love Montaigne felt for his friend, the fusion of their wills, rendered unnecessary the respect which Kant says should always be present in friendship. Montaigne tells us that he ‘banished from his thoughts’ all those words which implied separation and difference: ‘benefit, obligation, gratitude, request, thanks, and the like.’ 91

We should add that Montaigne did not think that this kind of friendship was common. On the contrary, ‘so many circumstances are needed to build it up that it is something if fate achieves it once in three centuries.’ 92 Although Montaigne was aware that in everyday friendships, ‘one must go forward, bridle in hand, prudently and with precautions,’ he still retained a capacity to throw the bridle away and discard precautions when they were no longer

87 Montaigne, 1988. (First published 1580.)
88 Montaigne, 1988, p. 97.
89 Montaigne’s friendship bears out Aristotle’s view that only virtuous people are capable of perfect friendship. Had la Boétie been less virtuous, Montaigne could not have trusted him so well, and the two of them could not have been as united.
92 Montaigne, 1988, p. 92.
required. In la Boetie he had met someone with whom this was possible. Montaigne would not have been able to have this friendship with la Boetie had he restrained himself, kept himself at a distance, as Kant advocates. The possibility of this ‘sublime’ friendship would have passed him by.

What would Kant have to say about Montaigne’s friendship? We could expect a sceptical reply. Kant understands the kind of friendship about which Montaigne speaks to be only a romantic fantasy. Complete friendship, Kant says, is the hobby horse of ‘writers of romances’ and of ‘poetical moralists.’ Montaigne, according to the Kantian sceptic, is either exaggerating, perhaps seduced by the nostalgic eloquence of his own prose, or is downright self-deceived in thinking he had discovered a perfect unity in friendship.

Let us take this scepticism as far as it can plausibly go. It is true that Montaigne speaks in superlatives. He tells us that his friendship with la Boetie ‘has no model but itself and can only be compared to itself,’ and of how such a friendship ‘is beyond the imagination of anyone who has not tasted it.’ Their union, he says, was brought about by ‘some inexplicable power of destiny.’ It is not difficult to suppose an ambiguity in Montaigne’s friendship with la Boetie. The sceptic would say that the passionate language that Montaigne uses in the essay was driven not so much by friendship, as by an infatuation of sorts, a kind of erotic adoration. After all, such intensity of feeling is not typical of the Essays. The sceptic would go even further and say that the Essays in their entirety were inspired by Montaigne’s loneliness and grief after the death of his friend. According to this interpretation, Montaigne was seeking solace in ink and paper. By writing the Essays he was able to continue a conversation with the dead friend, and recover some of the pleasures of self-revelation which he had had with la Boetie.

---

94 DV, 470 and LE, 422. Nietzsche also voices a scepticism about true friendship, pointing out the need to measure one’s words. Friendships, says Nietzsche, ‘almost always depend upon the fact that two or three things are never said or even so much as touched upon: if these little boulders do start to roll, however, friendship follows after them and shatters.’ (Nietzsche, 1986, p. 148. First published 1878-1879.) There is a bitterness in Nietzsche’s scepticism which is absent in Kant.
96 Pierre Leschemelle has this view. Biographers Donald Frame and Hugo Friedrich also agree that the Essays were inspired by the friendship with la Boetie. (Leschemelle, 1994, pp. 49-50; Frame, 1965, pp. 81-84; Friedrich, 1991, p. 241.)
Biographers are divided over the nature of Montaigne’s affections. But let us suppose that the sceptical interpretation has some truth in it. Would that mean that Montaigne’s essay on friendship had nothing to teach us? No, that would not follow. The sceptical interpretation does not alter the fact that the relationship he had with la Boétie was a great good in his life, something he would not have wished to be without. Indeed, on this interpretation, the friendship was his inspiration, his muse. If Montaigne did write his *Essays* out of grief for his lost friend, such a fact would only serve to reinforce how important the need is for the contact and self-revelation found only in intimate friendship. The point is that both the relationship with la Boétie and the afflatus for his great literary work would have been denied Montaigne, had he, Montaigne, abided by Kant’s rule that ‘even the best of friends should not make themselves too familiar with each other.’

To understand Kant’s insistence on the importance of keeping a distance from others, we need to recall why respect is so important. To respect another is to acknowledge her as capable of choosing her ends, of acting on reasons. We cannot respect another other than by putting her at a distance. But in friendship the salient fact about others is not their rationality. It is something else, an attribute which can only be described as a *presence*.

It is understandable why Montaigne’s friendship with la Boétie was so important to Montaigne and why, when la Boétie died, Montaigne felt he was ‘only half the man’ he had been. In friendship we seek and rejoice in the presence of another being like ourselves. We seek communion and contact; we seek to share ourselves and our experiences. As such, we do not perceive others under the description of ‘rational’ or ‘autonomous.’ It is not that we see them as a-rational or ‘things.’ It is just that agency is not the focus of our attention.

The need for friendship is pressing and powerful. One final example of friendship taken from Alexander Dolgun’s remarkable book, *Alexander Dolgun’s story. An American in the gulag*, will complete this section.

Dolgun’s story traces his imprisonment and torture in Stalin’s Soviet Union in the 1940’s and 1950’s. As a twenty-two year old American citizen working in Moscow in 1948, Dolgun was accosted on the streets of Moscow

---

97 Leschemelle firmly rejects the homosexual hypothesis but cites scholars who maintain it. (Leschemelle, 1994, p. 56.) Frame and Friedrich do not even raise the issue, although their silence could partly be due to the earlier era in which they were writing.

98 This interpretation is highly speculative and an equally good case could be made against it, bearing in mind, amongst other things, Montaigne’s rampant womanising and his insistence that friendship ‘is a spiritual thing.’ (Montaigne, 1988, p. 94.)

99 DV, 470.

100 Montaigne, 1988, p. 103.
by the MGB, the Soviet secret police. He was taken to the notorious Lubyanka interrogation prison where he endured 18 months of isolation and torture, before being sentenced to twenty-five years’ hard labour. Imprisoned in Lubyanka, Dolgun devised exceptional survival strategies. In constant pain due to repeated beatings, and not being permitted to lie down following his all-night interrogations, Dolgun trained himself to snatch sleep during the one minute intervals between the guard’s surveillance through a peephole. At other times he would pace his cell, imagining as he did so, that he was walking the streets of Moscow, then across the Soviet Union to freedom.

Despite his determination to keep himself mentally intact throughout his imprisonment, Dolgun did not lose the capacity for emotional engagement with others. After a time he becomes aware of a prisoner in an adjacent cell. The other prisoner, despite the rules, makes contact with Dolgun by tapping on his toilet pipe. This communication with another human being sends Dolgun’s spirits soaring.

I was laughing out loud for joy. I had a companion! A fellow human being was next door, a fellow sufferer, someone to make common cause with, someone who would care and understand. ... My heart was beating with excitement.

The guard on duty warns them both that they will receive hard punishment if the tapping continues. Dolgun is undeterred. ‘I knew we could work it out,’ he says, ‘and I was ecstatic.’

The ensuing rudimentary communication between the two men – two taps for ‘good morning’, two taps for ‘I’m back from interrogation’ – lifts Dolgun’s morale to the point that he becomes convinced he will be able to survive his imprisonment. By repeatedly tapping in patterns, the other prisoner tries to teach him a code. Dolgun says

I came to love the invisible nameless being next door who greeted me and said farewell. ... I felt his teaching to be a form of moral support. ... I heard him telling me to keep my courage up, that I was doing fine, that I would get through, and that he cared for me. That did as much to keep me alive and sane as the sleep I had won myself under the hat. Either one without the other would have left me seriously deprived.

Dolgun, 1975, p. 84.

Dolgun, 1975, p. 88. Dolgun managed to convince the authorities to return his broad-brimmed hat to him. Wearing the hat, he was able to fool the guards into thinking he was still awake when they saw him perched on his stool in the cell.
For three months Dolgun is obsessed by breaking the code. The moment he returns from interrogation he takes out his pieces of broken matchstick, and tries to make sense of the patterns in the tapping. All the time he fears discovery by the guards. Despite the sleep deprivation, his bodily injuries, and the relentless abuse of his interrogators, Dolgun keeps at it. His fellow prisoner, risking his life, repeats the code day after day. Finally Dolgun cracks it. What victory! ‘I felt a pure rush of blood in my chest for the man next door,’ he writes. At last he and his friend can call each other by name and swap stories. At that night’s interrogation Dolgun cannot stop giggling.

Dolgun wanted contact with the other prisoner just for the sake of having contact; just for the companionship and the sharing of experiences; just for assurance that he was not alone. He was prepared to risk everything to secure this fellowship. Dolgun describes the prisoner next door as a ‘friend.’ Even in these extreme circumstances where one’s very survival is at stake, friendship is not only possible, but priceless.

Imagine Dolgun, alone, starving, sleep-deprived, beaten, and with no prospect of release or escape. He hears two taps on the cell wall and he is ecstatic. Why? Because another rational agent is in the adjacent cell? Because at last in Lubyanka Prison he can realise a tiny Kingdom of Ends? Detachment, self-containment, a proper distance from others – these were the last things Dolgun needed from his fellow inmate and the last things he wanted to be. Dolgun in his cell, desperately poring over broken matchsticks, has much in common with Montaigne, poring over his *Essays* day after day in the study he had dedicated to his dead friend.

We are now speaking in the language of ‘needs.’ Friendship is based on need; something Kant would say means it has no moral content. Simone Weil says that friendship based on necessity is a corruption of the friendship. ‘Friendship is a miracle by which a person consents to view from a certain distance, and without coming any nearer, the very being who is necessary to him as food.’ Weil holds here as an ideal of friendship a universal love, similar to the benevolence of which Kant speaks. But normal friendship, as opposed to generalised benevolence, is based on need, a fundamental need to affirm that we are not alone. Aristotle was right in saying that we are first and foremost social creatures. The need for contact drives us into friendship. That is the lesson to be learnt from Dolgun and Montaigne. ‘There is nothing for

---

103 Dolgun, 1975, p. 95.
104 Weil, 1951, p. 135.
which nature seems to have given us such a bent as for society,' says Montaigne. 'Of a perfect society friendship is the peak.'\footnote{Montaigne, 1988, p. 92.} It is our nature to seek friends.\footnote{No doubt it is an evolutionary trait that has survival value. However, this is a point I will not pursue in this thesis.}

### 3.5 Liars and murderers: the narrowness of Kantian ethics

The task of this chapter was to put Kantian ethics to the motivational test. This meant addressing the question, 'can the motive of duty plausibly provide a motivating reason for action?' I said at the beginning that the answer was no, not always. There were two reasons for this answer. Firstly, Kantian ethics cannot motivate acts of friendship. This is because friendship, being a relationship grounded in voluntary affection, requires motives quite different from Kantian-style duty. The previous three sections were devoted to showing the plausibility of this claim.

It is now time to consider the second reason why duty cannot always plausibly motivate all moral actions. It is this. Duty sometimes provides a reason which is too narrow in its scope. This section is devoted to exploring and defending this idea. This point may be addressed more briefly than the preceding objection.

When one's motivating reason is duty, one is liable not to have taken adequate account of consequences. In the previous chapter, in trying to distinguish between a consequentialist and a non-consequentialist approach to ethics, a distinction was made between an ethic of responsibility and an ethic of conviction. The difference between these two approaches lay in the scope of responsibility the agent has. In adopting an ethic of responsibility, the person sees that they are responsible for \textit{the way the world goes}, that is to say, for what \textit{happens} as a result of his or her actions. If, on the other hand, a person adopts an ethic of conviction, she understands herself as being responsible only for acting upon her convictions, independently of how the world may go as a consequence. She is guided by principle, rather than by anticipated outcomes.

There are two ways of adopting an ethic of conviction. There is a Kantian way and a consequentialist way. It is the Kantian-style convictions which properly contrast with an ethic of responsibility. But let us consider first, by way of a reminder, how the consequentialist could also claim to be acting upon his convictions. The \textit{indirect} consequentialist might claim that he
also has convictions. After all, like the Kantian, he acts upon principles such as truth-telling and respect for others. It is just that he does so because such principles are justified on consequentialist grounds. He has no attachment to the principles in and of themselves; he is attached to them because they promote the relevant neutral values. We saw that Philip Pettit uses this approach to show how the consequentialist can show respect for persons.\textsuperscript{107}

This way of justifying convictions is different from the Kantian justification. For Kant, principles are justified by recourse to the idea of an autonomous rational agent. We must act in such a way that our reason for acting could serve as a reason for all agents. We should respect others because, like ourselves, they are sources of reasons, and are therefore capable of choosing their own ends. The first principles of morality are discoverable just by examining what it is to be the kind of being who can choose and act upon reasons, as opposed to being swept along by the contingencies of nature, as animals are. Therefore, the end point of justification for Kantian ethics is rational agency, not what may or may not occur in the world, as it is for the consequentialist.

If we accept the indirect consequentialist’s story about justification of principles, then the distinction between an ethic of responsibility and an ethic of conviction collapses. We are left with a distinction simply between two ways of understanding an ethic of responsibility. That is to say, we can be responsible for the way the world goes as either a direct consequentialist or as an indirect consequentialist. I argued in the previous chapter that indirect consequentialism could not plausibly pass the motivational test. There must be scope for acting upon an ethic of conviction, at least some of the time. But I should now like to suggest that one should not act upon an ethic of conviction all the time, as required by Kantian ethics, because to do so fails to take adequate account of our worldly existence. It is an unworldly, monkish outlook that requires that the way the world goes should never figure in our moral motivations. That it is unworldly and monkish is not surprising. That is exactly what Kant wanted. He was seeking, after all, a foundation of certainty for ethics. He believed that ethical principles could be derived \textit{a priori}. There was to be no gleaning of our ethical responsibilities by turning to the ‘foul rag-and-bone shop’ of the world, in all its unpredictability and grubby uncertainty.

This idea that one is only responsible for acting upon the right maxim, regardless of how things may turn out as a consequence, has its limits. There comes a point, when the stakes are high enough, when consequences have a

\textsuperscript{107} See § 2.2.3.
legitimate place. Kant’s insistence that one is obliged always to tell the truth is a case in point. I refer here to his discussion of whether it would be wrong to lie to a murderer who asked us whether a friend of ours whom he is pursuing has taken refuge in our house. Yes, says Kant, we should tell the truth to the murderer. To be truthful in all declarations is ‘a sacred command of reason prescribing unconditionally, one not to be restricted by any conveniences.’ Kant could hardly put the point more forcefully. It will be worth our while to explore the reasons behind his uncompromising position.

Korsgaard says we can best understand Kant’s duty of truthfulness by drawing on the Formula of Humanity version of the categorical imperative. This formula says

So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.

The problem with deception, says Korsgaard, is that it treats others as a means to an end. When I lie to the murderer I am trying to get him to adopt a certain end; I am trying to influence his actions in a way to which he could not assent. It is important to note here the nature of the ‘could not.’ Whether someone can or cannot assent to our end is not determined by whether he likes the end or whether it conflicts with his own ends, says Korsgaard. It is determined by whether he is in a position to assent. When confronted by the murderer, continues Korsgaard, the efficacy of my lie depends upon his believing that I am telling the truth. The murderer could not assent to my telling a lie because the murderer does not know that that is what I am doing.

The test of whether I am treating someone as a mere means is whether he can assent to my way of acting. For assent to be a possibility certain conditions must be met. He must have knowledge of what is happening and some power over the proceedings. If these conditions are not met, the concept of assent does not apply. The Formula of Humanity makes coercion and deception the most basic forms of wrongdoing: deny another the possibility of assent, and you deny his agency. That is why we are obliged to tell the truth. And keep promises.

108 SRL.
109 SRL, 427.
110 G, 429.
111 Korsgaard, 1986b.
112 Herman (1993g) uses this argument to counter the claim that Kant’s ethics is ‘canonical deontology.’ Contrary to common opinion, she says, Kant does not assert the priority of
We can now see why we should tell the truth to the murderer. To do otherwise is to show disrespect for his agency. Perish the thought that I should treat the murderer – now sanguinely sharpening his knife before my very eyes – as a thing. It is a problem, admits Korsgaard. Kant’s moral philosophy leaves us powerless in the face of evil. It ‘sets a high ideal of conduct and tells us to live up to that ideal regardless of what other persons are doing.’ In referring to ideals, Korsgaard identifies the critical difference between Kantian ethics and consequentialism, the difference seized upon by Weber in his distinction between an ethic of conviction and an ethic of responsibility. The difference lies in how the two ethical approaches answer the question, ‘How should we respond to the wrong-doing of others?’ Do we admit the world is not ideal, but work towards making it so, compromising as we go? Or should we live as though the world were already ideal and thus make it so within our own tiny sphere? Should we hold ourselves responsible for the moral imperfections of others or not?

The great moral theories divide over this question. Kant made his answer plain, and nowhere more so than in his essay, ‘On a supposed right to lie on philanthropy.’ We should live as though we were already in a Kingdom of Ends. We are not responsible for the evil of others.

This conclusion is intolerable, says Korsgaard. Just as Baron found Kant’s stoical attitude to the friend ‘dreadful,’ Korsgaard finds Kant’s position on telling the truth to the murderer ‘grotesque.’ And rightly so. Here is her proposed solution. There is more than one way to be a good moral agent. The Formula of Humanity is not the only formula of the categorical imperative. Korsgaard argues that the Formula of Universal Law does not yield the same dire results as the Formula of Humanity. We recall that the Formula of Universal Law states that

I ought never to act except in such a way that I could also will that my maxim should become a universal law.

The question is, could one will that lying to a murderer become a universal law?

This much we know: in general, lying is wrong. If everybody were to lie, the efficacy of lying would be defeated. Lies are only worth the telling
Reasons based on duty

because people in general do not tell lies, and there is a presumption of honesty and trust. However, Korsgaard claims to have found a loophole in the case of lying to a murderer. That is to say, she claims that under the Formula of Universal Law, it turns out that, while lying in general is wrong, lying to murderers is permissible.

Suppose everyone were to lie to murderers at the door. What would the result be? There would not be a problem says Korsgaard; lying would still be efficacious. Even if the murderer knows that it is universal practice for people to lie to murderers, he will not be expecting you to lie. He will not be expecting it because he presumes that you do not know that he is a murderer. Therefore, says Korsgaard, the maxim of lying to murderers is not self-defeating when it becomes a universal law, unlike lying in general.

Korsgaard says her argument shows that Kant is mistaken in supposing the two Formulas are equivalent. The Formula of Universal Law permits lying to the murderer, whereas the Formula of Humanity absolutely prohibits it. Her answer to the problem of evil is to act according to the Formula of Universal Law when dealing with evil.

The Formulas of Humanity and the Kingdom of Ends will provide the ideal which governs our daily conduct. When dealing with evil circumstances we may depart from this ideal. In such cases, we can say that the Formula of Humanity is inapplicable because it is not designed for use when dealing with evil.

She has exploited the fact that the Formula of Universal Law appears less strict than the other Formula. Maxims are somewhat flexible; whether an action is universalizable depends upon how the maxim is formulated. There is no corresponding flexibility in the Formula of Humanity.

Korsgaard has made a nice attempt at making Kant a little less unworldly. The trouble is, this strategy of switching Formulas smacks of fiddling the books, or in this case, twiddling the maxims, so as to arrive at a predetermined outcome. In suggesting that different Formulas may be adopted at our convenience in this way, Korsgaard is actually advancing the idea that we should be consequentialists when the stakes are really high and Kantians at other times. She would deny such a conclusion, but how else are we to

---

117 Korsgaard, 1986b, p. 346. See also Langton, 1992, § 4. Langton accepts Korsgaard’s solution to the problem of evil, but does acknowledge that it endorses ‘a kind of consequentialism.’ (Langton, 1992, n. 46.)
interpret her admission that lying to murderers is ‘grotesque’? It is grotesque because of the dire consequences.

But we must question whether the Formula of Universal Law – even with the twiddling of maxims – really does permit lying to the murderer.

The Formula of Humanity tells us that it is wrong for me to lie because lying does not respect the humanity of the other person. According to this Formula, lying is wrong even if the other person is a murderer. We now have a reason as to why lying to the murderer would not pass the test given by the Formula of Universal Law. We could not will that our maxim become universal law because if everyone lied to murderers everyone would be showing disrespect for humanity.

Suppose that we have thoroughly understood the Formula of Humanity. We know that it is wrong to force someone, through violence, threat or deception, to adopt an end that is not his own. We know that certain kinds of actions are prohibited. In such a scenario there is no need for us also to apply the universalizability test. There is no need because we know that if a certain action is wrong because it treats someone merely as a means, then it will be universally wrong. Wrong actions cannot be universalized. It follows that if the Formula of Humanity prohibits lying to the murderer, then we already know the result which must be given by the Formula of Universal Law.

Korsgaard assumes that in cases of evil, the Formula of Universal Law takes priority over the Formula of Humanity. But this is a nonsense; each formula is a categorical imperative. As Kant himself stresses, there can be many hypothetical imperatives, but only one categorical imperative. If some categorical imperatives were more equal than others, as Korsgaard suggests, they could not all be categorical imperatives. We cannot be categorically required (by the Formula of Universal Law) to override that which is absolutely, categorically prohibited (by the Formula of Humanity). If an action is categorically prohibited by one formula, then that must be the end of the matter. The formulas must be understood as equivalent, as Kant says, or the very notion of a categorical imperative loses its grip.

Korsgaard is mistaken in thinking that the two Formulas are so independent of each other that we can pick and choose which of them to act upon when the moral going gets tough. But we should be sympathetic to her mistake. She is trying to avoid, as a Kant sympathiser, being backed into a corner wherein she has no option but to endorse Kant’s argument that it is wrong to lie to the murderer. Unfortunately for Korsgaard, such a corner is where Kant’s ethics inexorably leads her. And that is what is wrong with Kant’s ethics.
The story about the murderer and the liar pushes Kant's theory to its limits and beyond. Kant, uncompromising in his belief that morality cannot be grounded in empirical considerations, will go where his theory takes him. We can now see the extreme position entailed by Kant's belief that 'the moral worth of an action does not lie in the effect expected from it.' \(^{118}\)

In reply to Kant, we would plead that the world in which we live is an imperfect one, that we cannot live in isolation entirely from the evil actions of others. Others' actions need to be anticipated by us or we run the danger of having too narrow an outlook, and permitting too much harm to be done to ourselves and others. It is true that by anticipating the murderer's actions and telling a lie we are permitting contingent circumstances into our deliberations, but in doing so we are only acknowledging the entirely relevant fact that we are situated in a world full of contingencies and chance.

As for Korsgaard, she wants to have her cake and eat it too. She wants to be a true Kantian, but she also wants to take responsibility for other people's evil. But the only way to have both a cake and the pleasures of eating it is to have more than one cake. To avoid the difficulties that Kantian ethics has in the face of evil, we must openly acknowledge the limitations of Kant's morality and adopt a pluralist outlook. That is, we must acknowledge that while Kant's ethics can motivate some of the time, it cannot motivate all of the time.

**Conclusion**

The arguments in this chapter have taken many twists and turns. Let us now try to summarise them. This chapter set out to determine whether Kantian ethics could pass the motivational test described in the Introduction and Chapter 1. If it were to pass the test, then acting from duty, that is, performing an action (or actions) just because it is the right thing to do, would be a plausible motive. While it may be admitted that duty can motivate in some circumstances, it cannot provide a complete set of moral motives for two reasons. Firstly, duty is incompatible with the motives required for friendship. Secondly, when acting from duty we restrict the scope of our responsibility too much to be plausible in all cases. Expected outcomes or consequences of an action can legitimately provide reasons.

None of this is to deny that the vision of morality expressed by Kant explains much of our common sense understanding about morality. It is

\(^{118}\) G, 401.
particularly relevant to explaining public morality and how one should view strangers or others to whom one does not stand in a personal relationship. This is so because it tells us that, regardless of circumstances, others are owed respect. We do not need to know others personally, like them, or have any other desires affecting them to know what our moral attitude to them should be. Kant was concerned to establish the freedom and equality of individuals and reject the pre-enlightenment belief that some individuals had more moral authority than others. In this regard, he has been immensely influential in the development of contemporary theories of liberalism and justice. John Rawls, for one, has taken much from Kant, by way of notions of self-governance, autonomy and respect for others.\footnote{See particularly Rawls, 1986, 1999b, 1999c. Also, as mentioned in the previous chapter, Benn’s theory of freedom has Kantian origins. (Benn, 1976, 1988.) Thomas Scanlon’s contractualism is also arguably grounded in Kantian-style respect for others: we should take seriously whether the contracting parties would accept or reject various proposed actions or courses of actions because the parties are sources of reasons. As autonomous reasoners their objections should be respected. (Scanlon, 1982 and 1998.)}

But what is appropriate for justice is not necessarily appropriate for friendship. The main argument in this chapter has been that one cannot act from Kantian duty and, at the same time, have the motives necessary for friendship. This is so because the affection and altruistic emotions which ground friendship, not only have no moral worth, but are not permitted the full expression which is required by close personal attachment. The moral emotions such as benevolence are to be encouraged only insofar as they support the motive of duty, and no further. And the pathological emotions, without which there could be no friendship, are to be so tightly controlled as to all but rule friendship out.

We have considered objection after objection which could be raised by the Kantian against this argument. At every tum we have tried to grant the Kantian as much as possible, but we saw that, on every occasion, there is a fundamental conflict between the requirements and outlook of duty, on the one hand, and the requirements and conditions of friendship, on the other hand. This conclusion should not be surprising. After all, it was not Kant’s aim to investigate friendship. His aim, as he saw it, was to investigate the nature of \textit{morality} which placed upon us demands quite independently of any inclinations or affections we may or may not have.

We saw that the Kantian could reply to the charge that duty and friendship are incompatible by saying that, while friendship has no moral worth because of its grounding in contingent emotion, it is still permissible so
Reasons based on duty

long as one does not violate one's moral obligations. A Kantian may have friendship, but only within certain limits. We replied to this line of argument by posing the question, 'But what kind of friendship would the serious Kantian be capable of?' Because one's emotional life would be so tightly prescribed and because one would always be required to place respect above all other attitudes, including love, one would only be capable of a very impoverished kind of friendship. To the extent that it is friendship at all, it would certainly not be the sort of friendship to which we would aspire and for which we would hope.

We conclude that we are not only choosing rational beings, as Kant thought. We are also feeling, emotional beings. Our capacity to be moved by others and our capacity to move others in return makes morality possible as much as our rationality, that is, our capacity to act on reasons. This is a fundamental point of difference between Hume and Kant's understanding of morality.

If we only see morality as residing in agency, certain morally significant features are denied. This, in the end, is why Kantian ethics cannot allow for friendship. We have to acknowledge that morality is made possible by a plurality of features, not just one.

Because of the belief that morality could not be grounded in anything which was only contingently good, it followed for Kant that the moral worth of actions could not lie in the outcomes they produced. We must look to the will behind the action, rather than the consequences which follow from the action, to determine whether the action is morally worthy. But this way of understanding moral action, as we saw in the last section, has its limits. There comes a point at which consequences have either so very much value or very much disvalue that they cannot be overlooked. In other words, there are situations in which we should understand that we are responsible for the way the world goes, not just for acting on convictions based on respect for others.

To understand the significance of the arguments against Kantian ethics put forward in this chapter we need to consider the arguments of the previous chapter and the anticipated arguments of the following chapter. The previous chapter addressed the question of whether consequentialism could provide motivating reasons. Here it was found that it, like Kantian ethics, could not motivate acts of friendship. Furthermore, it could not motivate a respect for people for their own sakes. In this chapter we find that respecting people for their own sakes is exactly what Kantian ethics motivates. However, in filling a motivational gap left by consequentialism, a new gap is created. Kantian ethics cannot properly motivate reasons based upon consequences. It does not
adequately allow for a responsiveness to the contingencies of the world in the way that consequentialism does. While consequentialism sees the worth of actions lying in their contingent relationship to the promotion of neutral value, and is too much dictated to by circumstance and instrumental considerations, Kantian ethics has the opposite problem. It does not take enough account of the effects of actions. The consequentialist has too broad a scope of responsibility while the Kantian has too narrow a scope.

It is now clear that consequentialism and Kantian ethics have at least one thing in common: an inability to motivate friendship. This is commonly understood to be the case because of the impartialist nature of both theories. Consequentialism requires that the friend be seen as another way of promoting neutral value, thus undermining the capacity for friendship itself. Kantian ethics require that we see friends in the light of what duty requires. Both of these approaches deny the intensely personal and affectionate nature of friendship. In the following chapter I consider another approach to ethics which can motivate friendship, but in the process, creates yet another motivational gap. While being able to motivate friendship, an ethic of care cannot motivate actions which require an impartialist outlook.
Reasons based on care

There is no necessity, that a generous action, barely mentioned in an old history or remote gazette, should communicate any strong feelings of applause and admiration. ... Bring (it) nearer, by our acquaintance or connexion with the persons, or even by an eloquent recital of the case; our hearts are immediately caught, our sympathy enlivened, and our cool approbation converted into the warmest sentiments of friendship and regard.

An enquiry concerning the principles of morals
David Hume

CONSEQUENTIALISM and Kantian ethics, despite having significant differences, have this much in common: they cannot properly motivate acts of friendship. In the case of consequentialism the incompatibility with friendship arises because we are required to be motivated by agent-neutral considerations. In the case of Kantian ethics, as we have just seen, friendship is excluded because acting from duty excludes the more particularised and affectionate motives necessary for friendship. Both theories are able to motivate an outlook whereby the agent sees herself as one player among many, giving no undue preference to either herself or to those she favours. In the process, such theories cannot motivate the outlook required for personal preferential relationships.

In this chapter I consider an ethical approach which can plausibly motivate friendship, but which has other limitations. I refer here to what is known as an ‘ethic of care.’ Whereas Kantian ethics places supreme ethical significance on individual agency and autonomy, stressing the necessity of each person acting upon reasons which are universalizable, the care ethic, in
contrast, emphasises the importance of responding to individuals in particular contexts and stresses the moral significance of relationship. It takes as its starting point the idea that we are all enmeshed within a web of relationships which requires care and attention for sustenance and flourishing. It has been claimed that an ethic based on sustaining relationships, through receptivity and responsiveness to others, is a characteristically feminine ethic. This emphasis on caring for others, a care which depends on the cultivation of sensitivity, appropriate emotional attitude and sound judgement, is a provocative development in contemporary moral theory. The ethic of care with which I am concerned in this chapter is driven in large part by feminists who observe that caring for others has traditionally been the role of women both within the family and the wider community and has been undervalued as a moral activity. It is, then, an ethical approach which emerges from a close look at women’s experiences and their ways of thinking about moral situations.¹

I argue that, were one to be motivated by considerations of care alone, one would have too narrow a moral outlook. One would place excessive emphasis on immediate relationships to the exclusion of those beyond the reach of one’s emotional engagement. The idea is that an ethic of care is parochial in its scope, giving inadequate moral status or guidance for action towards those removed from our immediate circle. Consequentialism and Kantianism are well able to motivate action towards those with whom we have no personal contact because both require the agent to adopt an impartial outlook, to stand back, as it were, and be motivated by principles and agent-neutral outcomes. Generally speaking, the more personal the relationship with others, the less appropriate it is to adopt an agent-neutral perspective. The more impersonal our relationship with others, the more it becomes necessary to have our relations governed by principles. The ethic of care cannot motivate action where principles and impartiality are required.

In the first section I identify some key features of an ethic of care, drawing on both empirical and conceptual understandings of care. I then go on to show in the second section that the motivational set provided by an ethic of care is too narrow and that, therefore, it does not properly pass the motivational test.

Before proceeding with these arguments I should acknowledge that the choice of the third theory being the ethic of care may seem surprising. There is a school of thought which says that the theoretical approach which stands in obvious contrast to consequentialism and Kantian ethics is virtue ethics. I do not subscribe to this view. Rather, I agree with Martha Nussbaum that understanding virtue ethics as the same kind of theory as consequentialism and Kantian ethics is to make a category mistake. Why this should be so is discussed in detail in the last section of Chapter 5. Also in that section I discuss a view of the virtues which arises from the arguments of the thesis in total.

4.1 Features of an ethic of care

Referred to largely by feminists wishing to challenge moral theory based on principles and rules, the ethic of care remains under-theorised. One problem is that it is not clear whether the ethic is intended to provide a descriptive or prescriptive account of moral activity. Authors slide between, on the one hand, the descriptive work provided by Carol Gilligan, who conducted empirical research into women's moral development, and on the other hand, prescriptive judgements which suggest its superiority to the prevailing ethic of justice. Another problem one faces in presenting an account of the main features of the care ethic is that there is no definitive account of it. Indeed, the approach itself is still evolving. Much work in the area is found in scattered papers in journals and edited collections and it is difficult to formulate a uniform position.

In this section I give a brief account of both the empirical research behind the care ethic, drawing on the work of Gilligan, and the more theoretical normative work in the area, drawing largely on Nel Noddings’s book, Caring: a feminine approach to ethics and moral education. Although Noddings is by no means the only writer working in this area, I focus here on her work because hers is one of the more comprehensive and widely cited attempts at articulating a theory of care. In the second part of this section I suggest that considerations

---

3 Nussbaum, 1999, p. 165.
4 On the vagueness and ambiguity of the idea of care, see Kuhse, 1995.
5 Noddings, 1984.
6 Another recent attempt is provided by Peta Bowden, 1997. Bowden's approach differs from that of Noddings in that she does not think that there is one account of care which
Chapter 4

of care can often plausibly motivate our actions towards others. The remainder of the chapter is devoted to showing that an ethic of care cannot always be a plausible motive of ethical activity.

4.1.1 Caring as an ethical activity

Empirical research by Gilligan and others suggests that women formulate moral problems and conduct moral reasoning differently from men. Gilligan distinguishes between an ethic of justice – which is based on principles of right conduct, and which stresses the separateness of individuals and their rights to non-interference – with an ethic of care, which stresses the nurturing of ties to others. The ethic of justice is based upon Lawrence Kohlberg’s theory of moral development which, following Kant, prizes the idea of acting on principle. In contrasting these two ethical approaches, Gilligan and Jane Attanucci write:

A justice perspective draws attention to problems of inequality and oppression and holds up an ideal of reciprocity and equal respect. A care perspective draws attention to problems of detachment or abandonment and holds up an ideal of attention and response to need. Two moral injunctions – not to treat others unfairly and not to turn away from someone in need – capture these different concerns. ... A care perspective involves the question of how to act responsively and protect vulnerability in a particular situation.

Gilligan argues that Kohlberg’s widely-accepted theory of moral development, which sees the highest level of moral development as being a stage in which one formulates and applies abstract universal moral principles, is biased in favour of male moral reasoning because it reflects the justice ethic. Based on a Kantian-style belief in the importance of individual autonomy, Kohlberg’s theory is that there are six stages of moral development. As the individual develops, their moral reasoning becomes more abstract. At the final sixth stage, according to Kohlberg, the individual has reached moral maturity. This maturity is marked by an ability to apply abstract reasoning to particular

---

8 Gilligan and Attanucci, 1988, pp. 73-74.
situations. It means being able to see which moral principle is appropriate and to give reasoned justification for action in terms of moral principles. Kohlberg’s description of this stage is remarkably Kantian. ‘The perspective is that of any rational individual recognising the nature of morality or the basic moral premise of respect for other persons as ends, not means.’

On Kohlberg’s scale women were often found to develop to around Stage 3, what Kohlberg calls the ‘conventional level.’ At Stage 3 the individual sees herself in relationship to other individuals. Shared feelings and the expectations of others outweigh individual interests. Rightness is understood as showing concern about others, ‘maintaining trust, loyalty, respect, and gratitude.’ But what distinguishes this stage from full moral maturity is the absence of a ‘generalised “system” perspective.’ Again, at this stage the individual is not properly mature because she wants to be good only to please others and to live up to their expectations. Gilligan theorised that women’s moral maturity did not follow the same path of development as that proposed by Kohlberg. It was not that women were immature because they frequently valued relationships over abstract principles. Rather, a different model of moral maturity was required. She called it an ‘ethic of care.’

Gilligan found that the care perspective is a distinctly female phenomenon. ‘It is notable that if women were eliminated from the research sample, care focus in moral reasoning would virtually disappear.’ The male conception of morality – the justice ethic – gives ‘lip service to intimacy and care but stresses at their expense the importance of autonomous judgement and action.’ The female perspective, on the other hand – the care ethic – stresses the ‘central importance in adult life of the connection between self and other, the universality of the need for compassion and care.’ Summarising Gilligan, Joan Tronto identifies three characteristics of the ethic of care. Firstly, it revolves around responsibility and relationships rather than rights and rules. Secondly, morality is tied to concrete circumstances rather than being formal and abstract. Thirdly, it is best expressed as activity, not as principles. The ethic of care stresses the particularity of each situation and does not search for the appropriate principle to provide moral guidance.

Gilligan's research, which was conducted by questioning volunteers about their moral decision-making, has been picked up by feminists keen to develop the idea that conventional moral and political theory — which stresses the need to respect the autonomy of others — is masculine, not gender-neutral, as it is purported to be.\(^{17}\) One difficulty Gilligan had in formulating the findings of her research was that, at the time, no normative ethical theory matched her findings. She says, "as a moral perspective, care is less well elaborated, and there is no ready vocabulary in moral theory to describe its terms."\(^{18}\) This was so, she supposed, because the concepts and language of moral development and moral theory were biased in favour of men's and boys' experiences and understandings. It should be noted, however, that this bias was by no means intentional. On the contrary, Kohlberg supposed that his theory of moral development was gender-neutral. Writing some thirteen years after the publication of her ground-breaking book, \textit{In a different voice}, Gilligan reflects:

> It was startling then to discover that women for the most part were not included in research on psychological development or, when included, were marginalised or interpreted within a theoretical bias where the child and the adult were assumed to be male and the male was taken as the norm.\(^{19}\)

It is not readily apparent what implications Gilligan's research, if statistically significant, has for a \textit{normative} moral theory. Her research simply made claims about how women \textit{actually} conceptualise ethical situations and work through the demands placed on them by those situations. By itself, it could make no claim as to whether women (or men, for that matter) \textit{should} think in this way, that is, whether it was desirable or laudable. These latter

---

\(^{17}\) It should be noted that Gilligan's methodology and findings have not been without critics. Reviewing the relevant empirical studies, Lawrence Walker disputes that there are significant sex differences in moral reasoning. He claims such differences are revealed only in a minority of studies, and 'even in those studies the differences tend to be small.' (Walker, 1993, p. 174.) See also Luria, 1993, and Greene and Maccoby, 1993, for a critique of Gilligan's methodology. The latter authors claim (p. 194) that her work attacks a 'straw man.' For a defence of Gilligan's studies see Baumrind, 1993, and Gilligan, 1993. Whatever the merits of Gilligan's methodology, however, her conclusions about different types of moral reasoning strike a chord of recognition with many women, including her critics. At any rate, the idea that there are two ways of understanding ethical maturity is interesting and plausible, whether or not these two ways parallel gender differences.


\(^{19}\) Gilligan, 1995, p. 121.
questions would need to be addressed in a normative theory. I turn now to consider the attempt by Noddings to develop such a theory.

To begin, any theory of care would have to answer the question: what is caring? Noddings says this is a difficult one to answer because there is no formula for caring which is applicable to each situation. The specific actions involved in caring for another depend upon ‘a constellation of conditions’ that is viewed through the eyes of both the one who cares and the one who is cared for.\(^{20}\) The main idea, however, is that of ascertainment and responding to need. This we do by apprehending the reality of the other: to the extent that we are able, we remove ourselves from the situation at hand so that we can focus on the other. According to Noddings, this attitude or outlook we have towards another is of great ethical significance. ‘An ethic of caring locates morality primarily in the pre-act consciousness’ of the one who is caring.\(^{21}\) The idea here is that in caring one is present to another. This means putting one’s own interests aside and adopting a posture of engrossment in the other. ‘Whatever she does for the cared-for is embedded in a relationship that reveals itself as engrossment and in an attitude that warms and comforts the cared-for.’\(^{22}\) This responsiveness and receptivity to another translates into action. However the actions will not be rule-bound and may well be different on different occasions since no two situations are exactly the same in all their detail. Caring is the activity of paying close attention to precisely the detail of the situation. She says:

\[
\text{Variation is to be expected if the one claiming to care really cares, for her engrossment is in the variable and never fully understood other, in a particular set of circumstances.}\(^{23}\)
\]

Such an ethic of care exhorts us to act in a way that serves to cultivate and sustain caring relationships with those around us. A caring relationship is one in which each party ascertains the needs of the other and, as far as reasonably possible, sets about meeting those needs. There can be no formula which can tell us in advance what, for any situation, would count as a caring response. Each person, as carer, must develop, using experience and example, the sensitivity and nous to perceive what each situation demands of her.

\(^{21}\) Noddings, 1984, p. 28.
\(^{22}\) Noddings, 1984, p. 19.
She rejects the idea of universal caring, that is, that we are obligated to care for everyone, 'on the grounds that it is impossible to actualise and leads us to substitute abstract problem solving and mere talk for genuine caring.' 24 We should care, then, for those most closely associated with our personal life because they are the ones we know most intimately and for whom we can take practical caring action. On the question of our obligation to care, we see a point of view in sharp contrast to Kantian morality. Whereas Kant understands moral obligation as arising from pure rationality, unsullied by the impurities of our worldly existence – experience, desires, sentiments and the like – Noddings sees our moral obligation as arising precisely from experience. Our obligation arises out of our experiences of being cared for and our finding ourselves in a situation of relatedness to others. We value the relation of caring which arises 'as a product of actual caring and being cared for and my reflection on the goodness of these concrete caring situations.' 25

She is sceptical also of institutional caring, saying that when institutions or agencies are established there is 'a shift of focus from the cared-for to the 'problem’ ... rules are formulated and the characteristic variation in response to the needs of the cared-for may fade away.' 26 It is the element of impersonality and rule-following, so common within institutional settings, which is corrupting of a caring ethical response. In this regard, Noddings’s approach is quite radical: she rejects totally the idea of a morality based on principles. Indeed, moral principles are opposed (on principle!) because they require us to abstract away from concrete situations. 27 The care ethic requires us to determine what response is appropriate to each situation, depending on a range of factors particular to that situation. She cites the example of lying to the school about her son being ill so that he can stay home to do something they both regard as worthwhile. Because her lying is conducted in the name of care, there is no need, she says, to justify her action by recourse to principles.

If I attempt to justify every disobedience or rejection of principle, I tacitly acknowledge that principles are paramount in ethical life. I shall turn the argument about and ask how he (the one asking for a justification of lying) might justify hurting his son by telling the truth. 28

24 Noddings, 1984, p. 18.
25 Noddings, 1984, p. 84.
27 Noddings, 1984, p. 5.
28 Noddings, 1984, p. 57.
I hope to have provided some indication of what is referred to as an ethic of care. It is important to be mindful that there are descriptive accounts and normative accounts and that the latter are based on the former. In the remainder of this chapter I am largely concerned with the normative point of view. Before moving on, it is useful to point out the appeal of a care-based approach to ethics.

4.1.2 Care as a plausible motive

The ethic of care is important in that it gives moral credit to activities typically undertaken by women. Simple acts aimed at maintaining and nurturing relationships, once seen from the perspective of the care ethic, may now be understood as ethical activities. It is all too easy to dismiss such mundane activities as being solely social in nature, part of a woman's or wife's or daughter's or mother’s role, and ethically insignificant. Perhaps in the past we have been tempted to perceive the moral as being grandiose, acting from principle, or acting courageously in public political forums, rather than in the small detail, the relentless domestic life of a woman. It is not suggested, however, that the ethic of care is or should be an ethical outlook which applies only to women. Indeed, for our purposes the issue of gender differences may be put to one side.

There is much that is right about an ethic of care. Both consequentialism and Kantian ethics have difficulties in motivating acts of friendship. An ethic of care, on the other hand, has no such problems. With its emphasis on the particularity of others and on the altruistic emotions of empathy, sympathy and benevolence, it provides a set of motives which the other two theories cannot. It awakens us to the minutiae of ethical activity in close relationships, which was strikingly absent in the two previous theories. But as with consequentialism and Kantian ethics, it cannot plausibly motivate all ethical action. If we were only motivated by our care for others in the way described by Noddings, the ambit of our ethical lives would be too narrow. The ethic of care cannot adequately reach those people beyond our immediate circle and contact.

4.2. The limits of care as a motive

In this section I examine the question of whether the ethic of care could plausibly motivate. I argue that, like consequentialism and Kantian ethics, it
could not, on its own, provide a complete motivational set for ethical action. As I have already suggested, the set of motives which it provides is too narrow in its scope. It could only plausibly motivate within a very narrow circle of interest. In the first part of this section I show why it is too narrow. Then later in the chapter, to illustrate its narrowness, I consider a care-based ethic in the context of nursing. An ethic of care has been used as the basis of a professional nursing ethic, but I argue that, were nurses actually to be motivated by it, they would not be adequately able to respond to a range of issues that arise in a nursing context.

4.2.1 The narrow ambit of care

What would the motivational set be like of people who adopted the ethic of care? They would be too narrow in their outlook. Two points arise in relation to this narrowness. Firstly, the care perspective accords moral status only to those with whom we have a close relationship. Secondly, it does not give us the concepts and language to respond morally to states of affairs beyond our immediate context. Let us consider each problem in turn.

(i) The attribution of moral status

The ethic of care stresses the importance of relationship, of responding to those around us in an attentive way. What is wrong with this? In the case of consequentialism we noted earlier that the world would surely be a better place if we were all more assiduously motivated by the consequences of our actions. So, too, in the case of ethics of care, surely the world would be a better place if we were more receptive to the needs of those around us, if we cared more about others. To put the problem simply, a care-based ethic leaves too many people out of the ethical equation. Those close to us do not exhaust those who are, and who should be, morally significant for us. By restricting our moral world to those of whom we have a personal knowledge, the ethic of care promotes arbitrariness in moral response.

To see how arbitrariness arises, consider the following example. In times of disaster when thousands of people may be adversely affected by earthquake, flood, famine or war, the media frequently focus our attention on the plight of one or two individuals, perhaps a child trapped at the bottom of a building under the earthquake rubble. Many people show a keen interest in this individual's welfare. The media provide regular updates, and many find themselves hoping for the victim's well-being, perhaps donating money or
other items to their cause. This phenomenon shows at once the strength and limitation of the care ethic. It is not enough to explain away what is going on here by reference to media manipulation and the drive for a ‘good story.’ That the reporting of the story may be other than morally motivated does not detract from the genuineness of the readers', listeners' or viewers' response. If such stories do generate a lot of interest it is because we are prone to feel great sympathy with the suffering of another, particularly an innocent such as a child, once their situation is made known to us. In this case the media connects us through a ‘chain of caring’ to the stranger whom we may never meet in a country which we may never visit.²⁹ Our response is the hallmark of a set of moral sentiments which lie at the heart of an ethic of caring.

But it is clear that, although this response may be heartening, it is very limited. What makes it so limited is that there are many others who are equally worthy of our interest and attention but who are unable to make a claim to it. Although a similar plight may be shared by numerous others, yet we remain unmoved. The interest generated by the child trapped by rubble arises largely because the situation is perceived as being both individual and close to us, in just the way Hume describes in the quotation opening this chapter. Our inability to care in a direct personal way for large numbers of people means that all too often our attention becomes consumed by an individual in need and broader questions which address the needs of the unknown others are left either unasked or ignored.

There are practical limits to the degree to which we are able to attend to others as individuals. We cannot give to all and sundry the same attention which we give those close to us. What is more, those bonded by friendship and kinship are entitled to more. As Martha Nussbaum says

If, for example, we could ever see clearly and be moved by the value of each unique person in the world, we could never without intolerable pain and guilt be able to act so as to benefit any one of them rather than any other - as love, or justice might require. (If I saw and valued other people's children as I do my own, my own could never receive from me that love, time and care that she ought to have, that it is just and right for her to have.)³⁰

²⁹ ‘Chains of caring’ is a phrase used by Noddings to show how we may be linked to unknown individuals. She has in mind chains which link those we don’t know ‘to those already anchored in the inner circles.’ (Noddings, 1984, p. 47.) However, the media can also link us to strangers.

For one thing, we do not have the emotional capacity to respond to large numbers of people with the same intensity with which we are able to respond to a single individual. It requires a great deal of time and attention to appreciate fully another’s individuality. Our caring capacity, where caring is understood in this direct, immediate way, is finite. So we take shortcuts. We abstract away from others’ particularity. The further removed we are from people, the more abstract becomes our relationship to them and the less their individuality features. The closer the connection with others, the more we relate to and appreciate their individuality; the more distant the connection, the more abstract the connection becomes, and the more principles need to be invoked. We do not need to know about the personal qualities and idiosyncrasies of another to know that, for example, we should respect their autonomy.

We can now see how the charge of arbitrariness can be sustained against an ethic of care. The ethic of care, as expressed by Noddings, encourages and obliges us to respond only to a chosen few. To be sure, one would not wish to dispute that we have relationships deserving of close attention and a caring response in the case of family members, friends and neighbours. But in the case of those beyond our immediate circle, we care for those who arbitrarily appear before us. In cases of catastrophe, the chosen few may be selected by a media whose interests lie in sensationalising suffering. In other cases, it is a matter of chance, rather than need or desert or moral worth, as to who happens to cross our path. Our moral attention is directed arbitrarily.

It is to an ethic of justice, rather than an ethic of care, that we need to turn when there is suffering on a grand scale. An ethic of justice can accord moral status to those who are otherwise only statistics precisely because it does not depend on we, the moral agents, having a close personal caring relationship with them. We are connected, by way of moral principles, to those of whom we do not have direct knowledge. In the case of the disaster or the famine an ethic of justice gives us the moral language and conceptual understanding to ask: how many in total are injured or displaced or hungry, and what resources are required to meet their needs? Who is obliged to provide the resources? What measures can be put in place to ensure such an occurrence is avoided in the future? How can aid be better distributed? An ethic of principle enables us to consider the moral obligations, based on notions of justice or fairness, of our own government and community to contribute to long-term solutions to chronic problems of poverty, famine and the like.

Furthermore, because of its exclusivist approach to morality, the ethic of care lends itself to an attitude of, at best, indifference and, at worst, antipathy and hostility towards the excluded. It promotes a ‘garden gate’
approach to moral thinking and action, whereby our moral world extends only as far as the gate at the bottom of the garden. Those on the other side of the gate, the passing strangers, or those in places removed from our garden, have no moral significance for us. If these others have no moral significance is it not a small step then to adopt an exploitative, scapegoat, domineering or intolerant attitude towards them? Such a step is prone to occur in respect of nationalistic sentiments. When the garden gate extends as far as the national borders, we identify with our nation and compatriots to the exclusion of others. It then becomes easy to blame other nations in times of hardship or adopt an aggressive posture towards them. Acts of aggression can be more easily justified and a warring attitude more likely. Such an attitude arises from an ‘us and them’ understanding of moral obligation. Those with whom we identify we accord moral status and those with whom we do not identify we accord no moral status.

Such malevolence towards the ‘others,’ those who are not cared for within the confines of our moral garden, is not a necessary consequence of the ethic of care by any means, any more than extreme nationalism is a necessary consequence of pride in one's national community. But, I suggest, if the care ethic is not tempered by notions of justice which enable us to include in our moral community those with whom we do not have a personal link, such malevolence is more easily adopted because such an ethic divides people into two categories, those for whom we care and those for whom we do not care.

At any rate, even if it were the case that the care ethic did not encourage negative behaviour towards those removed from us, such behaviour is not morally prohibited. In fact, because those writing about the ethic of care focus only on the positive behaviour of caring, it is not specified what our attitude should be towards those with whom we do not have a caring relationship. There is no talk, for example, of negative obligations, such as simply not interfering with the non-cared for and it seems unlikely that such an obligation would be endorsed, since it is a derivative of the libertarian idea of the right to non-interference, that is, a key notion in what may be regarded as a justice ethic. The care ethic wishes to distance and contrast itself from the approach of the justice ethic.

Noddings is explicit that there is no need even to justify lying if it is done in the name of caring for those close to one.\footnote{Noddings, 1984, p. 57.} One simply does what one judges necessary to meet the needs of the one who is cared for. What if one judged necessary the harming or deprivation of another in the interests of the
one cared for? What are the limits of caring? If there are no limits the care ethic results in a siege mentality with everyone 'out for what they can get' on behalf of those they care for. If, on the other hand, there are limits how can these limits be set without recourse to moral principles?

(ii) The absence of social criticism

The second problem about the ethic of care is that, because of its narrow scope, it does not enable us to give a moral critique of wider society. The care ethic is restricted to a small circle of people, namely those with whom we can have a close relationship. We have no obligation to consider those beyond our small circle. The narrow ambit of the ethic of care means that it cannot properly support or even properly articulate the interests of, say, women as a group or section of the community. To give a critique of a gendered society at all requires the abstract language of justice. As Mary Dietz points out, feminists have not had a critique in the language of care but in the language of politics.32 The care ethic has nothing to say about oppression or inequalities between the sexes. The most it might say about power is that those who exercise it have needs and they should be cared for by those close to them (presumably women, since Noddings refers to carers in the feminine pronoun).

On the one hand, the ethic of care stresses the importance of recognising the interconnectedness and interdependency of individuals. It understands us as living within a network of relationships which sustain us and which we, in turn, are obliged to sustain, the opposite of the liberal understanding of people as autonomous agents who intentionally choose their relationships. In this sense the ethic of care appears to be a non-individualistic ethic. On the other hand, ironically, this same ethic promotes individualism. This is so because it does not enable us to make generalisations about others. It perceives people as unique individuals with specific needs which vary from context to context. In doing so, it denies the perspective whereby one may be conceptualised as being a member of a wider group. Such a conceptualisation would be criticised as an attempt to abstract away from the particularities of an individual's circumstances. The advocate of care would claim: we must respond to each individual and their particular circumstances as we are confronted by them. But abstractions and generalisations are necessary to some degree if we are to provide a critique of our situation, as a member of a group, within the wider society. The care ethic does not lend itself to making generalisations about a

group in the community, in this case, women, and the needs of that group as a collective.

It might be objected that the needs one has by virtue of being a group member can all be taken into account by the care ethic. Here is how. When I see another before me I take into account, in deciding how best to care for her, her group memberships: her gender, her class, race, cultural background and so on. Whatever general properties the group has I apply to her as an individual. In this way the care ethic can incorporate the situation of women as a group.

This line of reasoning may go some way towards meeting the objection, but it does not go far enough. First, as already mentioned, the care ethic would have to abandon its dislike of abstractions and generalisations. It would have to concede that it is not adequate to view each individual as unique in their circumstances and needs, and that it is possible to ascertain another’s needs simply by having an awareness of her group memberships.

Second, even with such a modification the care ethic would still be concerned with individuals responding to other individuals. Some issues concerning women and their interests can only be properly addressed at a level of principle. Consider the following example. One issue which continues to appear on feminists’ agenda has to do with the idea of ‘equal pay for equal work.’ Suppose we had a situation of a group of men and a group of women working at the same level of responsibility, the same number of hours and the same number of tasks. Suppose the men were paid twice as much as the women. Since all other things were equal, the only conclusion we could draw was that the women were paid less on account of their simply being women. What would the care ethic have to say about this scenario? It would not even have the language with which to criticise it because such criticism would have to appeal to the principle of equality between the genders. Such principles are excluded because they necessitate abstracting away from the concrete example.

One response we might make according to the care ethic perspective might be this: the woman feels angry at her unfair treatment (and she would only be justified in feeling so if we accepted that there are moral principles which should come to play in such matters). We, as a friend or colleague or family member, care about her and so we listen attentively to her complaints. We sympathetically suggest she have a soothing massage. More than that, we generously lend her our favourite CD of New Age relaxation music to calm her nerves. In other words, we try to behave in a generally sensitive, caring manner.

---

33 The same is true for issues concerning any identifiable group which claims it is being treated inequitably, such as the disabled and indigenous people. Even the idea of inequity requires abstraction.
We can only respond to the fact that the woman we care for feels angry or
anxious. We cannot respond to the reason for the anxiety, namely, the injustice
of the situation because the care ethic cannot even articulate that the situation is
unjust. The care ethic is limited precisely because it operates only at the level
of particular individuals’ needs. It rejects moral principle, including those
principles of justice and fairness which are required to support, amongst other
things, feminist criticisms and analyses of current practices and theories.

It might be objected that if everyone cared for those around them then
that would be morally sufficient. But even this much is not clear. We can
imagine a community in which everyone, for example, has close family ties and
everyone is assured of being cared for in times of ill health or difficulty. But it
is still conceivable that such a community has no political autonomy and is at
the mercy of other communities. We can imagine a society based on classes
where one class dominates the others, yet everyone within the classes cares for
others within that class and indeed cares for some in other classes. We could
imagine a society in which people are attentive to each other as individuals, but
in which there are structured inequalities in terms of wealth, services, freedoms
and opportunities.

In summary, the characteristic which is mooted as the great strength of
the ethic of care – its priority on the tangible, the particular – is precisely its
limitation. It confines us to the here-and-now, and to those about whom we
have specific knowledge. No consideration is given to those beyond our circle,
including unknown future generations. Secondly, the ethic of care, on its own,
does not give us the language or moral understanding to describe and criticise
whole social and political landscapes. As I noted in the first part of this
section, the care ethic relates only to individuals. It does not relate to groups
within communities or to comparisons between communities. Questions to do
with inequalities cannot be addressed. Noddings refers to the starving children
in Africa, saying we have no obligations to care for them since they are so
removed from us.34 Perhaps we are not personally obligated to go to Africa to
care for those in need. But do we want to surrender the moral language with
which to describe and explain unjust arrangements? It renders us unable to pass
moral judgement on situations which are not immediate to us. If we were
motivated by the ethic of care, the starving in Africa would be none of our
moral business. Our moral business would have to do only with those nearby.

34 Noddings, 1984, p. 86.
4.2.2 The ethic of care as a professional ethic

In this the final sub-section of Chapter 4, I would like to give an example of how the motives provided by an ethic of care would be insufficient in a professional context. Various authors have used the idea of care, both to explain what the practice of nursing involves, and to articulate a nursing ethic.\textsuperscript{35} However, the ethic of care, as articulated along the lines given by Noddings, is limited in this context. Furthermore, it would be similarly limited for other professions which involve caring for others, such as teaching, social work, counselling and the practice of medicine in its various branches. All the ethical situations which are likely to be encountered by such professionals working within an institutional setting cannot be captured by an ethic of care. In other words, the care ethic, on its own, provides an incomplete account of nursing ethics.

One way in which nursing care differs from care given to those with whom we have a close personal relationship, for example, family members and friendships, is that nursing occurs within a more public setting. I especially have in mind here nursing within an institutional setting. On this point, Bowden writes ‘the context of nursing care is typically more formally organised, its practices more directly regulated by external forces’ than the context, say, of maternal care. Therefore, nursing practice is subject to ‘the determinations of publicly administered norms and structured by the demands of publicly sanctioned conduct.’\textsuperscript{36} Nurses work with at least two sets of demands upon them, those of the institution – I would include here demands by doctors as well as administrators – and secondly, those of the patient. For this reason Hugo Engelhardt refers to nurses as ‘in-between’ people.\textsuperscript{37} Nurses, therefore, are not engaged in the activity of caring for others’ health sim[pl]iciter; they care for others’ health within the constraints of an institutional setting. This setting, which brings nurses’ caring role into a public domain, requires that ethical considerations more aligned with an ethic of justice be brought into focus.

We have already observed that understanding ethics solely in terms of care means that ethics becomes confined to a personal private domain. Ethics of care cannot capture ethical issues which arise at a more personally removed, public level. To repeat a point made earlier, the further we move away from

\textsuperscript{35} See, for example, Benner, 1984; Benner and Tanner, 1987; Carper, 1979; Fry, 1989; Leininger, 1984 and Watson, 1985.
\textsuperscript{36} Bowden, 1997, p. 101.
individual personal ties, the more abstract our ethical thinking needs to become. If we do not abstract at all, those with whom we have no personal tie, and about whom we have no concrete, specific knowledge, are omitted altogether from the ethical equation: they end up having no moral status. The situation of nurses is particularly interesting in this context because they inhabit a world which has elements of both personal care and personal relationship, as evidenced by the way they relate to particular patients, and elements of the public world of morality, where decisions need to be articulated and defended, and resources ought to be allocated according to some notions of fairness.

Insofar as nurses inhabit the first world, the ethic of care is highly relevant. It enables us to theorise about what nurses actually do in their relationships with individual patients, and, in terms of the motivational test, what could plausibly motivate their relations with them. Insofar as they inhabit the second world, they are more subject to considerations which belong to an ethic of justice. For example, nurses and administrators all have to make decisions about the allocation of resources. In a world where financial considerations come to bear on all medical practice, difficult decisions must be made as to which areas have priority for funding. At this point, a concept of justice is relevant. There are simply not the resources to ensure that every patient receives the kind of attention and equipment they would receive in the best of all possible worlds, where there was no constraint on funding. Although these decisions about the allocation of resources are not always the responsibility of nurses, nonetheless nurses work within constraints dictated by the fact of limited resources, and need to decide, for example, how much time and attention they give to individual patients. How much care can be delivered to how many patients? To answer this question, some general principles are required of the kind: ‘nurses have an obligation not to neglect any patient’ and ‘time must be equitably divided among patients.’ Without such principles, we would be at a loss to know how practices involving caring can be properly and justly arranged. The ethic of care can tell us a great deal about the importance and nature of caring, particularly in one-to-one situations, but to make decisions about priorities among large numbers of individuals requires frameworks which can only be delivered through recourse to principles.

Another way in which ethical principles come to bear in nursing contexts is in the justification of decisions. Principles enable us to bring discussion of ethical matters into the public domain. The ethic of care, with its emphasis on the carer’s response, places a great onus on the carer having the ‘right’ sensitivity, that is, a perception and response which is appropriate and relevant to the situation. In the context of more private relationships such as
friendship, the need to justify one’s responses and persuade others that they should respond similarly, is usually not so pressing. When caring becomes more public, as it does in the case of professional carers, there is greater accountability for behaviour and decisions. Carers need to be able to defend decisions and points of view, and this defence involves engaging in some kind of principled reasoning. Where there is a dispute between nurses, both nurses may be able to claim that they are attentive and sensitive to the particularities of the case, but unless they are able to abstract away from those particularities and show why they are ethically significant, they are unlikely to be able to resolve the situation. As Helga Kuhse says, ‘as far as ethical discourse is concerned, she will be condemned to silence.’

Kuhse makes the point that the ethic of care provides an incomplete account of ethical activity, in all contexts, not just that of nursing, because it denies that there is any role for principles at all. She says that the dichotomy between ‘concreteness’ and ‘abstraction,’ which mirrors the distinction between ethics of care and ethics of justice, is a false one.

Principles need contextual concreteness; and contextual concreteness needs abstraction ... a moral approach always requires some abstraction from the total context ... the question is not whether context is relevant, but rather what elements of that context ought to be ‘abstracted’ from the overall context as significant for ethical decision-making.

Kuhse’s point is well taken. If we had no principles whatever, how would we be able to sift through the myriad of features and aspects to every situation we find ourselves in? How would we know which features were significant and which we should ignore? In the context of nursing, how would nurses be able to justify decisions? Let us agree that Kuhse is right in saying that ethics of care and ethics of principle differ in degree and not in kind. It will still remain true that certain situations, namely those, I suggest, which are located in the private domain, closer to the self, demand that much greater weight be given to contextual variation than situations which demand a response to people outside the private domain, whom we have not met. Within institutions where decisions have to be made regarding large numbers of people, again, ethics of principle are required. Nursing is an interesting example of an activity which demands close personal attention, engrossment of the kind Noddings speaks of,

---

but which exists within a broader context of justice. This broader context provides a limit to the engrossment of the carer. The ethic of care, with its disdain of principles, cannot by itself address issues arising from this broader context, such as allocation of resources and accountability for carers’ decisions.

Conclusion

Unlike consequentialists and Kantians, not all those advocating an ethic of care claim that the idea of caring captures everything which may be thought of as being in the moral domain. Gilligan maintains that an ethic of care and an ethic of justice provide two complementary ways of understanding ethics. There was never any suggestion, she says, that an ethic of care could usurp the place of an ethic of justice.40 Similarly, in writing about the altruistic emotions of compassion, benevolence and sympathy, so necessary for friendship, Blum stresses that his challenge to the Kantian view is one of showing the limits of such a view. His argument is not that there is no place for acting on principles of obligation; rather, he argues that ‘it is, often, morally good to act from universal principles, or from a sense of duty.’ His point is that such actions do not ‘exhaust the area of morally good action.’41 We should be motivated by different kinds of reasons according to circumstance.

We saw in earlier chapters that both consequentialism and Kantian-style ethics of principle had difficulty accounting for the motives appropriate to close relationships such as friendship. In this chapter we have seen that a third ethical approach provides motives entirely appropriate to friendship and other one-to-one relationships, but at the price of not being able to motivate in other contexts. The particularism implicit in ethics of care is, at one and the same time, its great strength and weakness. •

41 Blum, 1980, p. 93.
A plurality of reasons

To every thing there is a season, and a time to every purpose under the heaven ...

Ecclesiastes 3: 1

HAVING now arrived at the final chapter, we would do well to summarise the story so far. In the first chapter we established, by way of the motivational test, that a moral theory must be able to provide plausible motivational reasons. Each of the following three chapters then went on to consider a moral theory with a view to ascertaining whether or not the theory could plausibly provide such reasons. These three theories were consequentialism, Kantian ethics and the ethic of care. In each case, the set of motivations provided by the theory was unsatisfactory because of other motivations that it ruled out. More specifically, each theory contained motivational gaps which could only be plugged by the other two theories.

There is merit in the consequentialist perspective. There is a place in one’s moral thinking to see the world from an agent-neutral perspective, a place for stepping back from a self-interested or parochial outlook and thinking carefully about the impact one’s actions and dispositions will have upon the world at large. However, it is a mistake to have all of one’s actions motivated by such outcomes. We saw that important kinds of actions are excluded by a consequentialist set of motivations. In friendship, we value others for their own sake, just for who they are. Acts of friendship lose their distinctive friendly character if they are motivated by agent-neutral outcomes. Similarly, the idea of respect for others is grounded in the worth of individuals as ends in themselves, not in the outcomes that showing respect has on the world at large. Respecting
others means that we do not see them just as a means to bringing about some desired outcome, even a desirable agent-neutral outcome. Individuals matter for their own sakes. And certain actions towards others should be motivated by a recognition of that point.

A Kantian outlook however, while being capable of motivating respect for persons and thus being capable of filling one of the gaps left by the consequentialist perspective, also has motivational blind spots. One who was motivated by Kantian respect and duty would be incapable of friendship, though for different reasons from the consequentialist. Friendships are personal relationships sustained by emotional attachment. They require motives antithetical to duty. Furthermore, the Kantian, committed as he or she is to acting upon principles based on respect for others, would be too little motivated by the contingent outcomes of his or her actions. This latter motivational gap, of course, can be plugged by consequentialism; the consequentialist is only motivated by contingent outcomes. The third possibility, an ethic of care, can provide a set of motives appropriate to friendship and can therefore accommodate a problem of both the consequentialist and Kantian, but not without its own blind spots. The ethic of care, with its emphasis on the moral emotions directed at particular individuals, is too narrow in its scope. It has too little to say, unlike Kantian ethics and consequentialism, about the moral status of those we do not know or with whom we have little in common.

No one of the three theories has it all. If we are motivated by impartialism and disinterestedness, we miss out on partialist relationships. If we are motivated in such a way as to be able to sustain partialist relationships, we miss out on disinterestedness. If we are motivated by outcomes, we fail to see others in a non-instrumental way. If we see them in a non-instrumental way, we fail to be sufficiently worldly. The kinds of considerations which can and ought to motivate us are, quite simply, too many and too varied in kind to be captured by any of the three theories discussed.

The purpose of this final chapter is to draw out the implications of all this. In the first section I suggest that the preceding four chapters, when taken collectively, amount to an argument for pluralism. I submit that there are multiple sources of moral reasons. It is as simple and as complicated as that. Also in this first section I defend the selection of the three theories discussed in the preceding chapters. The second section considers various supposed problems with pluralism, including the problem of conflict. I suggest this problem can be best dealt with by recourse to certain personal qualities or virtues of the agent faced with the situation. In other words, the answer is to be
found in virtue ethics. Accordingly, the third and final section considers the view of the virtues which arises from the pluralism which is being defended. In this last section I also return to the question, raised at the beginning of the last chapter, as to why virtue theory was not considered as the third theory to be contrasted with consequentialist and Kantian theories.

5.1 An argument for pluralism

The motivational test has shown us that no single kind of consideration could plausibly motivate moral action. All three moral theories considered in previous chapters can motivate some actions but their motivational capacity has limits; none can explain the full range of moral motivations. We are left to conclude that, in our practical deliberations and theoretical understanding of moral matters, there is a plurality of moral reasons. We thus find ourselves in agreement with Charles Larmore who, in arguing for the heterogeneity of morality, cites consequentialism, deontology and partiality as the three principles of practical reason.1 In consequentialism, deontology and care we find the three most fundamental principles to which we are morally subject.

It still remains for the choice of the three theories to be defended; the three approaches to ethics are not the only ones available. I have not, for example, discussed ideal spectator theories or contractualism or virtue ethics. These latter three approaches, it might be said, cover much of moral philosophy’s landscape. In this section I argue that these other candidate theories are not to be considered as rival theories because for them to offer substantive principles about right action they need to refer back to one or a combination of the three theories already discussed. (For easy citation, the three theories discussed in Chapters 2, 3 and 4 will henceforth be referred to collectively as ‘the three theories.’ Ideal spectator theories, contractualism and virtue ethics will be referred to as ‘privileged authority theories’ for reasons which will shortly become clear.)

Take Thomas Scanlon’s contractualism.2 Scanlon holds that principles of conduct must be able to be justified to reasonable co-operative agreement makers. He says that

1 Larmore, 1987, Ch. 6. The principle of partiality ‘requires that we show an overriding concern for the interests of those who stand to us in some particular relation of affection.’ (p. 132.)
An act is wrong if its performance under the circumstances would be disallowed by any system of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced general agreement.3

Our actions must be justifiable to others. What is able to be reasonably rejected can only be determined by reflecting on the sorts of reasons 'to which we would give relevance and weight in thinking about what co-operative life with others requires.'4

It is hard to quarrel with the idea that one should be able to justify one's actions to others, particularly when those others are reasonable and committed to social co-operation. The point at issue is not whether it is true that right actions are those which could reasonably be accepted by others. The point at issue, rather, is whether actions are right because they could reasonably be accepted by others. At the start of the first chapter we referred to the question posed in The Euthyphro: do the gods love something because it is holy, or is something holy because the gods love it? We saw that the question is all about whether there are independent standards which justify certain responses or judgements. An analogous question can be asked in respect of the privileged authority theories. The Euthyphro question applied to contractualism asks: is an option right because it is justifiable to co-operative agreement makers, or is an option justifiable to co-operative agreement makers because it is right?5 Scanlon maintains the former answer: what makes an option right is that it can be justified to the relevant others.

But it remains open to us to ask why the contracting parties would see fit to reject or approve a particular principle. Or, to put it another way, by virtue of what is a rejection deemed reasonable rather than unreasonable? In deciding, for example, that φ-ing is to be rejected, the parties would need to give reasons. Those reasons would be along the lines of 'φ-ing is disrespectful to others' or 'φ-ing is harmful to others' or 'φ-ing overlooks the importance of personal attachments' or something similar. The hypothetical parties to whom the principle is presented must have reasons for rejecting it and it is those reasons which make the principle wrong, not the mere fact that it was rejected.

---

5 This argument against the privileged position theories is similar to that given by Pettit against such theories. (Pettit, 1997, pp. 136-140; 1999 and 2000b, pp. 159 ff.)
These reasons, in turn, would be grounded in one or a combination of the three theories.⁶

To bring this point home, let us consider the question from the point of view of the participating ‘others.’ Suppose I am one of the hypothetical group of reasonable others and I have to decide whether to reject φ-ing. How do I arrive at my decision? I do so by appealing to factors about φ-ing such as those indicated above. Is it disrespectful? Is it harmful? And so on. If there is disagreement in the group, presumably we will discuss the issues as we each see them. We will appeal to various principles or the context in which the φ-ing will occur. I am not permitted simply to reject φ-ing arbitrarily or on a whim.⁷

Scanlon says that what is basic to his approach is ‘the idea of justifiability to each person (on grounds that he or she could not reasonably reject).’⁸ To be sure, he admits that those others must have reasons for accepting or rejecting an option. But Scanlon does not equate wrongness with the reasons the option is unjustifiable; he equates it with the unjustifiability itself. His refusal to ground wrongness in any antecedently identifiable principles makes his account of wrongness and rightness circular in the way that all analytic truths are circular.

Here is an analogy. Suppose you are a novice chess player. Wanting to improve your game, you ask a Grandmaster what you need to do to win. The Master replies, ‘move your pieces in such a way that your opponent’s King is unable to move without being in check.’ You would understandably be dissatisfied with this reply. Certainly the Master has told you something true, but has he been helpful? No. His answer contains only an analytic truth about what it means to win a chess game. You already knew that. You wanted a

---

⁶ The parties might also reject a principle on the grounds of their own self-interest. It might be that one individual, though committed to finding agreement, would refuse to accept a principle because it required him to make excessive self-sacrifice. But reasonable self-interest could be grounds for rejection. The contractualist approach acknowledges that each party to the contract is a source of reasons; that each person’s reasons count. Respect for persons in a way articulated by Kant is arguably the justificatory basis of contractualist approaches. The Kantian influence on John Rawls’s contractualism is very evident and has already been noted in Chapter 3, n. 119.

⁷ Pettit makes a similar point. The contracting parties, he says, argue with each other about the various alternatives on the basis of how those alternatives fare in regard to certain values. Because the parties’ reasons must ultimately refer back to values, says Pettit, questions about rightness should turn to value theory for answers. For him the best value theory is consequentialist in kind. (Pettit, 1997, pp. 136-140.) It should be noted, however, that Scanlon is not concerned with actual agreement, only hypothetical agreement. (Scanlon, 1998, p. 155.) But it still remains the case that the hypothetical others must have (hypothetical) reasons for their agreement or disagreement.

⁸ Scanlon, 1998, Ch. 5, n. 8; see also p. 154.
substantive answer about what you should do to place your opponent’s King in inescapable check. Saying that right action consists in action which is justifiable to others is similar to the Master’s answer. Even if it were true, it would not be giving any substantive guidance to rightful action.

To opt for the first choice of the *Euthyphro* question is always to opt for an analytic style of answer to a question of judgement. It is to say, ‘that which is holy, good, right, desirable and so on is that which is valued or desired or chosen by A, where A is some authority. All the privileged authority theories make this claim. Where the theories differ is in explaining who A is and what gives A moral authority. In the case of contractualism, A’s authority is derived from the fact that A is a co-operative agreement maker. In the case of ideal spectator theories, A is a person with an idealised set of desires or a person with all relevant information and flawless reasoning. In the case of virtue theories, as we shall see, the authority lies in the fact that A is a virtuous person. While there is genuine disagreement about what should ground A’s authority, all approaches agree that rightness is determined with respect to A’s authority. In the same way, the first choice of the original *Euthyphro* question says that holiness is determined by reference to what those in undisputed authority – the gods – love.

The contractualist, like all advocates of privileged authority theories, faces a dilemma. Either the contractualist opts for the first possibility of the *Euthyphro* question, or he opts for the second. If he opts for the first – an option is right because it is justifiable to co-operative agreement makers – contractualism only offers an analytic truth about rightness. But if he were to opt for the second – an option is justifiable because it is right – he would have to admit to there being antecedent principles of rightness, and this would take him back to the three theories, or versions thereof. Either the contractualist is offering us an analytic truth about rightness, or he must draw on the substantive principles offered by one or a combination of the three theories. We conclude that, though the contractualist may well be right in saying that rightful action is justifiable to others, he is not offering any new substantive principles of rightness. For this reason contractualism, and the other privileged authority theories are not considered as alternatives to the three theories.

As we draw this section to a close, let us consider briefly ideal spectator theories. Just as the contracting parties would have reasons for their judgements, so too would the ideal spectator. Ostensibly the ideal spectator’s choices and judgements are right because she is placed under ideal conditions of deliberation. The idea is that if the conditions are ideal, meaning she has the right desires and beliefs, full information, and the like, she will grind out the
right decisions. But what makes the ideal spectator's judgement so impeccable is that such a one knows how to compare the various demands placed on her; when to be a consequentialist, when not; when to place the obligations of personal relationships above those of impartial morality, and when not to, and so on.\(^9\) Not only are the conditions for making judgements ideal, but so is her judgement itself. She is motivated in the right way by the reasons provided by the three theories. The ideal spectator theorist, like the contractualist, offers no new substantive moral principles.

The third of the privileged authority theories, virtue ethics, is discussed in the last section. It will be argued there that, like ideal spectator and contractualist theories, it too faces a dilemma. Either virtue theory offers only an analytic truth about rightness, or it must appeal back to one or a combination of the three theories for substantive moral content.

One point still needs to be made about the three theories. I have not defended the idea that there are only three theories that offer substantive moral principles. For that to be true the three sets of motivations collectively provided by the three theories must comprise a complete set of moral motivations. I have not proved that this is so. All I have done is show that the motivational blind spot of each theory in respect of the areas I specify is picked up by one of the other theories. That is, there is no need to appeal to any other theory beyond the three discussed to provide a complete motivational set in respect of the areas discussed. The three theories complement each other nicely. But it is always possible that there remain further areas which I have not discussed, in respect of which all three theories contain motivational blind spots. Even if this were so, however, it would not affect my claim of pluralism.

### 5.2 Problems for pluralism

A pluralistic approach to ethics is not new. Charles Larmore, Thomas Nagel, William (W. D.) Ross, Michael Stocker and Bernard Williams would all agree that we should abandon the search for a systematic and unified moral theory and recognise a variety of distinct moral considerations that are not linked by

\(^9\) As McNaughton says, it has sometimes been thought, incorrectly in his opinion, that the ideal spectator would be utilitarian, having an impartial interest in the happiness of each. (McNaughton, 1988, pp. 166-167.) However, if the arguments of the previous chapters are correct, the impartial spectator would be pluralist, not monist.
common principles.\textsuperscript{10} In this section I defend the conclusion of pluralism against several objections which may be levelled against it.

\subsection*{5.2.1 The problem of simplicity}

Let us begin with the problem that is most easily dismissed. When it comes to moral theories, Philip Pettit says a premium should be placed on simplicity. He says it is ‘common practice in the sciences and in intellectual disciplines generally to prefer the more simple hypothesis to the less, when otherwise they are equally satisfactory.’\textsuperscript{11} Consequentialism, he says, fares well in this regard, being ‘indisputably a simpler hypothesis’ than non-consequentialism. This is so because the consequentialist only has one response to value, whereas the non-consequentialist has two responses.\textsuperscript{12} The consequentialist must always \textit{promote} value, whereas the non-consequentialist will sometimes promote value and sometimes honour it. Non-consequentialists could agree, for example, that values such as economic prosperity and the safety of nuclear installations should be promoted, rather than honoured. Consequentialism wins on the grounds of simplicity because it only requires one response to value. One response is always simpler than two.

Now simplicity can certainly be a great advantage, not only in intellectual disciplines, but in many walks of life. For example, advertising copywriters would be well advised to keep their slogans simple. In a world where intricate and many-faceted messages need to be communicated via a thirty second sound byte, simplicity is essential. ‘Read my lips: no new taxes’ is a simple and memorable campaign slogan for a would-be President-elect. The question is, is the simplicity of consequentialism akin to the desirable simplicity of good scientific theories or is it more like the facile simplicity advanced by marketing gurus? That is to say, does the simplicity reflect a simple reality or is it a Procrustean simplicity? If the arguments of the previous chapters have been successful we can only conclude that it is the latter, Procrustean kind. By excluding friendship and non-instrumental respect for

\textsuperscript{10} Larmore, 1987, Ch. 6; Nagel, 1979c; Ross, 1955, pp. 18-30; Stocker, 1990, 1997 and Williams, 1981e, 1985a, pp. 100-101 and 111-113.

\textsuperscript{11} Pettit, 1991, p. 237.

\textsuperscript{12} On this occasion Pettit contrasts consequentialism with non-consequentialism according to \textit{responses} to value rather than \textit{kinds} of value. But the point could also be made in respect of kinds of value: the consequentialist says that only agent-neutral values are morally relevant, whereas the non-consequentialist says that both agent-relative and agent-neutral values are relevant.
A plurality of reasons

From our range of moral motivations, consequentialism pays too high a price for simplicity.

Pettit goes on to say that not only is non-consequentialism complex, but that it is complex in an *ad hoc* way. What is it about some values that they require honouring rather than promoting, he asks. Not only do non-consequentialists ‘have a duality then where consequentialists have a unity; they also have an unexplained duality.’13 But the non-consequentialist does not have an *ad hoc* division of values as Pettit suggests. There is no grand list of values into which the non-consequentialist, blindfolded and disoriented, sticks a pin, thus determining those which require honouring. We saw that the notion of respect can be derived from the idea of agency and autonomy. Friendship, on the other hand, is valued because of certain human needs. Both of these values can be traced back to two different aspects of our nature, rationality on the one hand, and social need on the other.

There is the added point that the non-consequentialist, in drawing attention to agent-relative values, reflects the way people are actually motivated. We saw that people do not value friendship in an agent-neutral way and that if they did there would be something peculiar about them. This alone is enough to defend the non-consequentialist against the charge of *ad hoc*-ery.

### 5.2.2 The problem of coherence

Whereas Pettit places a premium on simplicity, David Brink places a premium on coherence.

According to Brink, coherentism places a premium on systematic explanation. The more inter-connections and mutual support among beliefs, the better those beliefs are justified. Because of this, coherentism favours unified moral theories over fragmented moral theories.14 Pluralism is just such a fragmented theory; it says that moral considerations are irreducibly distinct and unrelated.

However, Brink is willing to concede the possibility that ‘The most coherent set of moral beliefs could be radically pluralistic or fragmented.’15 Brink’s coherentism is thus a *prima facie* consideration, like Pettit’s simplicity. The presumption in favour of unified moral theories could be defeated, says Brink, if it turned out that all unified moral theories ‘have as many defects as it is claimed utilitarian theories have, and if nonunified theories could account for

our considered moral beliefs. 16 He does not think that this is the case, although the possibility that it could be is reason enough to take seriously the objections to his favoured theory, objective consequentialism.

The thesis has argued that Brink’s condition is met. Utilitarian theories do have defects as do the other two theories.

A further problem remains. As Brink says, ‘By making moral considerations irreducibly distinct and unrelated, non-unified moral theories have a hard time explaining how to resolve moral conflicts.’ 17 The question of how to resolve conflicts between different moral considerations cannot be ignored. In the remainder of this section I consider this issue of conflict, arguing that the resolution of conflicts can be best understood by recourse to the intellectual virtue of judgement. The discussion of judgement will take us to a broader issue of the place of virtue within a pluralist system.

5.2.3 The problem of conflicts

‘The obvious difficulty facing a pluralist system’ observes David McNaughton ‘is the problem of moral conflict.’ 18 Where there is a plurality of moral principles or considerations there is always the possibility that more than one might apply in any given situation. However, as McNaughton goes on to point out, conflict may be a difficulty for pluralism but it is not itself an objection. On the contrary, being able to accommodate conflict can be seen as a strength of pluralism. 19

In this section I argue that, insofar as conflict is a problem, it is no new problem. Each of the three theories discussed already faces the problem of conflict within itself. Similarly, the solution – judgement – is no new concept.

(i) Conflicting principles

It might be thought that both consequentialism and Kantian ethics alike have the advantage of a system which does not face the problem of irresolvable moral dilemmas; there is always, in principle, one obligation which is overriding. 20 While Kant maintains that there can only ever be one overriding

---

18 McNaughton, 1988, p. 194.
19 McNaughton, 1988, pp. 194-195. We saw in the Introduction that Stocker argues for pluralism on the grounds that it can explain moral conflict.
obligation, he does admit that there may be conflicting grounds for a choice. But how is one to know in any given situation which of these grounds constitutes one’s (categorical) duty? It seems conceivable that there could be situations where the only way of keeping a promise is to lie. Suppose I promised my friend I would protect her from the murderer who is pursuing her. (At the time it seemed most unlikely that the murderer would be able to trace her to my residence; I am in the habit of covering my tracks.) The only way I can keep my promise when the murderer shows his face at my door is by lying. Should I keep my promise or tell a lie?

Ontologically, there is no conflict in such a situation; there is only one categorical duty. But there may well be an epistemological conflict. It is little help to someone facing this kind of situation to know that one or the other grounds of obligation takes priority in principle, but not know which one. The difference between the Kantian and the pluralist in respect of conflict is that the pluralist contends that conflict occurs at both the ontological and epistemological levels.

Similarly, it might be thought that a utilitarian can avoid conflict. Since happiness is the common currency, he can simply, ‘aggregate the pluses, subtract the minuses, and pursue the options with the highest balance.’ But even the utilitarian may face conflicts between different kinds of happiness. At any rate, consequentialists in general do not escape epistemological conflicts. We recall that consequentialists do not specify what the values to be promoted are. There could well be instances of conflict between different values, for example, social stability and freedom.

Joseph Butler was an epistemological pluralist but an ontological monist. We recall from Chapter 2 that he solved the problem of conflict between consequentialist and deontological considerations by appealing to God. God’s purpose in giving us our nature, supposed Butler, was to bring about the general happiness. Even though it often seems that acting from deontological considerations does not bring about the greatest overall welfare, this is only because our understanding is finite. As mere mortals, we lack the celestial perspective. If we had it we would realise that, despite appearances, everything always works out for the best. Although we cannot grasp how these seemingly disparate principles may be reconciled, we must have faith that they are. According to Butler, we must be content to trust in the ‘divine convergence’ between consequentialism and deontology.
Butler’s ‘divine convergence’ solved a difficult theological problem, namely, ‘if God wants us to bring about the general happiness, how come morality requires us so often to act in ways that do not bring about the general happiness?’ However, it did not solve the problem of the parishioner on the ground, so to speak, who still had to make a decision about which consideration was strongest at any given time. And that still remains the problem. The pluralism being discussed here is both epistemological and ontological. The three kinds of reasons are incommensurable, that is, they ‘cannot be precisely measured by some common scale of units of value.’ How do we know, then, which of the three sources of value ought to prevail in any given situation? We need to turn to certain faculties on the part of the agent. Butler told us to look to our consciences. He could not be more specific than that. Indeed, an answer can only be given at a general level. It is a matter of that most nebulous and ancient concept: judgement.

(ii) The importance of judgement

Contemporary moral philosophers have tended to downplay the significance of judgement, if not neglect it altogether. Historically, however, from Aristotle to Kant to Adam Smith, judgement has been recognised as having an important, if not central, place in moral theorising.

Judgement is referred to in two ways, descriptively and evaluatively. We can say, ‘Deciding between two options is a matter of judgement,’ or, ‘She showed judgement in choosing option A.’ In the latter case, what is meant is, ‘She showed good judgement.’ In the former case the descriptive sense is used. The question then is, what is judgement? If we know what judgement is we will have a better idea of what good judgement is.

Larmore’s description gives us a useful starting point. We exercise judgement, he says, by ‘responding with reasons to the particularity of a given situation.’ Two words provide focal points, ‘particularity,’ and ‘responding.’ Broadly speaking, there are two aspects to judgement, the what and the how. The what relates to the question of which features of a situation at hand are morally salient. This part of judgement relates to the ability to recognise morally relevant features and their relative significance. The how relates to the way in which one should respond in that situation. Judgement, therefore, tells us what we should respond to and how we should do it.

---

24 Chang, 1997a.
26 There may be more than two aspects to judgement; the point is there are at least two.
Those philosophers sympathetic to the ethic of care tend to stress the importance of the first aspect of judgement, moral perception. Indeed, any kind of particularism will tend to emphasise the importance of this aspect of judgement. If right or wrong action is determined on a case by case basis then the capacity to discern the relevant particulars of each case becomes critical. We must have a sensitivity to a situation so that 'no relevant feature escapes us, and we do not mistake its relevance either.' The more particularist the theory, the more work will need to be done by notions such as 'sensitivity' or, as McNaughton puts it, 'moral vision' and its metaphorical counterpart, 'moral blindness.'

But judgement is not just a necessary part of care ethics. It plays an important part in Kantian ethics, although it has not always been emphasised by those sympathetic to a Kantian approach to ethics. Consider what it is to act on principle. Believing that the principle is correct and having the moral commitment to act on it, is insufficient. One must also know how to apply it. The gap between believing that a principle is correct and applying it in a particular situation can only be bridged, says Kant, by judgement. Judgement is the faculty that links theory with practice.

Though understanding is capable of being instructed, and of being equipped with rules, judgement is a peculiar talent which can be practised only, and cannot be taught. It is the specific quality of so-called mother wit; and its lack no school can make good. ... A physician, a judge, or a ruler may have at command many excellent pathological, legal, or political rules, even to the degree that he may become a profound teacher of them, and yet, none the less, may easily stumble in their application. For, although admirable in understanding, he may be wanting in natural power of judgement. He may comprehend the universal in abstracto, and yet not be able to distinguish whether a case in concreto comes under it.

In applying a principle, judgement is again required in two ways. First, we need to judge whether the particular circumstances at hand are those to which the principle applies. Second, if we believe that the principle is relevant to the situation, we still have to judge the best way of applying the principle. For example, a principle, such as 'do not discriminate on the grounds of race,' does not itself tell us whether any particular situation is a case of such

27 See, for example, Blum, 1991 and 1994.
28 Dancy, 1993, p. 64.
29 McNaughton, 1988, p. 205.
30 KrV, Bi72-173.
discrimination. Nor does it tell us whether a particular course of action would fulfil the requirements of the principle. This knowledge about whether the situation at hand meets the criteria relevant for the application of the principle, is not something which is imparted to us by the principle itself. So even though morality is understood as acting upon universalizable maxims, one must apply judgement when formulating maxims.  

To implement the principle, ‘promote neutral value,’ even the consequentialist still requires judgement. It is a matter of judgement as to which value should be promoted in any given situation. For example, one may need to choose whether a particular situation calls for the promotion of the neutral value of freedom or the neutral value of peace.

Though Kant is clear in the passage cited above that judgement cannot be taught, being ‘a natural gift’ and a ‘natural power,’ elsewhere he says its acquisition can be facilitated by others.  

Kant says that teachers are to instruct their pupils in the analysis of every duty. They should ‘let the assembled children test their understanding by having each say how he would solve the tricky problem put to him.’ Although such discussion may facilitate the acquisition of judgement, practising and having exemplars are also important. The process of moral development takes us beyond the scope of the present topic, however. The relevant point is that the idea of judgement is already implicit in each of the three theories. It is not an alien concept to either the consequentialist, the Kantian or the care theorist, although its role may not always be explicitly acknowledged. The suggestion here is that there is an additional role for judgement in being sensitive to which of the three kinds of reasons should be decisive in any particular situation. This additional role is present in the first part of judgement, perceiving the relevant features of a situation. Salient features can be divided into three categories: consequentialist, deontological and care. Judgement is required to determine which of the three kinds of reasons is appropriate for any given situation.

Judgement may be thought of as an intellectual virtue. It is the virtue which integrates the three kinds of moral considerations. In suggesting it as the answer to the problem of conflict, I am following the thought Max Weber had in his essay, ‘The profession and vocation of politics.’ Weber believed the

---

31 Herman, 1993c, pp. 75-78.
32 DV, 477-484. Similarly, Aristotle, for whom judgement is critical for the exercise of the virtues, holds that it may be acquired and perfected through experience and habit. See Sherman, 1989, Ch. 5.
33 DV, 484.
34 See § 2.1.3 for discussion on Weber’s essay.
A plurality of reasons

politician faced an irreconcilable conflict between an ethic of conviction and an ethic of responsibility. Rather than finding a theoretical solution to the problem of conflict by appealing to a still higher principle, Weber appealed to the personal qualities of the politician. The way the politician responds to this conflict is a measure of his virtue as a politician. Weber says ‘the ethics of conviction and the ethics of responsibility are not absolute opposites. They are complementary to one another, and only in combination do they produce the true human being who is capable of having a ‘vocation for politics’.535 Weber thought the rational man recognised both kinds of demands, but more than that he was unable to say. Butler’s answer to the problem of conflict between consequentialist and deontological demands was similar. There was no higher principle to which he could appeal, so he turned to the inner capacity of the people who were faced with the problem. He invoked the notion of conscience. Likewise, appealing to judgement solves the problem of conflict in pluralism, to the extent that such a problem is solvable.

5.3 Pluralism and the virtues

Now that the three theories have been argued for, and the possibility of their giving rise to conflict has been acknowledged, it is time to consider questions of virtue and character. In this final section we turn to the virtues in general, plotting their relationship to the three theories already discussed.

There are three parts to this final section. The first of these explains why it would be a category mistake to understand virtue ethics as the third theory of the thesis. At the beginning of Chapter 4 I indicated that in this last chapter there would be a full explanation as to why virtue ethics was not to be taken as the third alternative theory to consequentialism and Kantian ethics. The first sub-section turns to that explanation. The second sub-section goes on to defend the view that virtue ethics is not an alternative to the three theories by showing that all the features which might be considered unique to virtue ethics can be accommodated by one or a combination of the three theories. Virtue ethics, in other words, does not offer us anything over and beyond what the three theories are able to provide. The final section goes on to outline a view of the virtues which falls out of the pluralistic position for which I have been arguing.

535 Weber, 1994, p. 368. (First published 1919.)
5.3.1 Virtue ethics and the three theories

The modern revival of virtue ethics may be traced back to Elizabeth Anscombe's 1958 paper, 'Modern moral philosophy.' Anscombe says that understanding ethics in terms of moral obligation is unhelpful, since the notions of 'obligation' and 'ought' only indicate the survival of concepts outside the framework of thought that really made them intelligible. That framework, she says, was provided by the idea of divine law, which had dominated in previous centuries. She exhorts us to think about the virtues instead. It would be a great improvement, she says, if, instead of talking about what is 'morally wrong' we were to talk about a genus such as 'untruthful,' 'unchaste,' and 'unjust.' Rather than describing an action as 'wrong' we should 'ask whether, e.g., it was unjust; and the answer would sometimes be clear at once.' In deciding whether to perform a certain action, you consider whether it is, say, an act of injustice and 'according as you decide it is or it isn't, you judge it to be a thing to do or not.'

There is now such a variety of authors writing on the subject from various angles that it is difficult to isolate one definitive account. Nonetheless, virtue ethics - with its emphasis on character, not just rightful action - is understood by many as being the rival theory to consequentialism and Kantian ethics.

Martha Nussbaum has recently suggested, however, that it is a category mistake to see virtue theory as contrasting with consequentialism and Kantian ethics, and I agree. This is so, she says, because these other theories also have an account of the virtues. This is a point to which we shall shortly return. Nussbaum suggests that this mistake arose for historical reasons. Prior to Anscombe's paper, there had been a dissatisfaction with the turn that utilitarianism and Kantian ethics had taken. The proponents of these theories were over-emphasising reason and choice at the expense of feeling and a life

36 Anscombe, 1958, pp. 5-6.
37 Anscombe, 1958, p. 9.
38 Anscombe, 1958, p. 12.
41 Nussbaum, 1999, p. 165.
A correction was badly needed, says Nussbaum, but it is not to be found by postulating virtue ethics as an alternative theory.

The same point to that raised in respect of ideal spectator theories and contractualism can be made in respect of virtue ethics. Either virtue ethics offers an analytic kind of truth about rightness, or it must appeal back to one or a combination of the three theories for substantive content about rightness.

To see that this is so, we need to ask what prescriptions virtue theory gives us. Rosalind Hursthouse tells us that virtue ethics yields the practical prescription, ‘“do what the virtuous agent (the one who is honest, charitable, just, etc.) would do in these circumstances.”’

The question now is, what *would* such an agent do? We can only reply that she would do the honest, charitable, just thing. Hursthouse says as much. She tell us that we cannot specify in advance what exactly the virtuous agent would do in any particular set of circumstances. We can only speak of ‘the sort of thing’ that agents would do and the ‘sort of truth’ that one would have to grasp to know, for example, that deception may not be a kindness. She says that if we could explain what is required by virtue any better, then virtue would amount simply to blind rule-following. It would be the kind of rule-following which ‘clever adolescents’ and ‘whiz-kids’ would be good at. But, says Hursthouse, virtuous action is not at all like clever rule-following.

This reply does not advance us one iota in our quest for substantive content to the notion of virtue. The argument is circular. Of virtue ethics we ask the question: what should we do? The reply is: do what the virtuous person would do. We ask: what would the virtuous person do? The answer is: act virtuously.

Hursthouse’s practical prescription has no substance to it. Virtuous acts are those performed by a virtuous person and there is nothing more to be said. The only way that we can give substance to the idea of a virtuous agent is to appeal to some antecedently identified features and to say that the virtuous person is sensitive and responsive to those features in the appropriate way.

---

44 Hursthouse, 1996, p. 30. A similar view is held by John McDowell, who says ‘a kind person has a reliable sensitivity to a certain sort of requirement which situations impose on behaviour. ... a kind person knows what it is like to be confronted with a requirement of kindness.’ (McDowell, 1979, pp. 331-332.) The point here is that we want an account of the requirements of kindness which does not refer back to a kind person.
To bring this point home, let us now turn to a couple of examples from Michael Slote, who advocates a common sense virtue ethics.\textsuperscript{46} Slote thinks that it is nothing unusual to appeal to virtues in moral problem solving. He says that the United States Vice-President, Al Gore, advocated the North American Free Trade Agreement (NAFTA) by appealing to the national character. Gore said that to reject NAFTA would be ‘to adopt a cringing, fearful, or despairing attitude to the world and America’s future.’\textsuperscript{47} There is nothing unreasonable about making such an appeal, says Slote.

There can be no doubting Slote’s point that the language of virtue has currency. We constantly judge ourselves and others in the language of virtues and vices. We refer to people as, ‘kind,’ ‘generous,’ ‘mean,’ ‘cowardly,’ ‘humble’ and so on. But how are we to resolve disputes about judgements of virtue? Suppose someone were to challenge Gore. ‘Mr Vice-President, Sir,’ says his opponent, ‘to reject NAFTA would be courageous. It would show leadership and pride in our American way of life.’

The only way the virtue and counter-virtue claims can be argued is by appealing to justificatory principles like justice or maximising welfare. The point is that appealing just to virtues is no moral argument on its own, any more than appealing just to ‘obligations’ or ‘duties’ is. We want reasons to underpin the virtue judgements. We want facts, principles and theoretical frameworks presented coherently and sensibly. If the virtues are understood in Hursthouse’s circular way, Gore would be required to say to his opponent: ‘Rejecting NAFTA would be to adopt a cringing, fearful attitude because only a nation with a cringing, fearful attitude would reject NAFTA.’ And what sort of reply is that?

To be sure, the appeal to virtues and vices has rhetorical value, as Gore was undoubtedly aware when he spoke of a ‘cringing, fearful, despairing’ attitude. Gore’s appeal tells us more about techniques of political persuasion than about moral problem solving. The language of virtue has tremendous rhetorical force. Nobody wants to be fearful, cringing or callous.\textsuperscript{48} It is a truism to say that most, if not all, people want to be — or in a more Machiavellian

\textsuperscript{46} Slote, 1997.
\textsuperscript{47} Slote, 1997, p. 231.
\textsuperscript{48} As an aside, we observed in the first chapter (§ 1.4.1(i)) that Cora Diamond uses the language of virtues when criticising Peter Singer. His arguments are not ‘right,’ ‘wrong,’ or ‘problematic;’ they are ‘shallow’ and ‘obtuse.’ The idea is that this kind of epithet gives a more damaging and more profound criticism. But as Nussbaum properly points out, in the end we need to talk about substantive views. (Nussbaum, 1999, p. 201.) Diamond’s virtue-talk only has substance (as opposed to rhetorical value) once her opponents’ arguments have been laid out and exposed for flaws.
spirit, want to be seen to be – decent, respectable people. It is easy to refer to the virtues when trying to persuade or present an ideal. But the virtues only become philosophically interesting when we look behind them to see what grounds them.

Slote’s other example refers to someone having to decide whether to prolong long and painful treatment for her mother’s terminal illness. The mother is too ill to decide for herself. What the daughter should do, says Slote, is inform herself as much as possible about her mother’s condition. Then, assuming the facts are fairly clear cut and point to ‘horrendously painful and debilitating prospects’ for her mother, the woman’s decision to withdraw treatment is derivable from considerations of benevolence. ‘At that point, it would be callous of her to insist on heroic measures and benevolent not to do so and the proper moral decision can thus be reached by agent-based considerations.’ There would be nothing strange or unusual about the daughter justifying her decision by reasoning that ‘it would be ... callous of me to try to keep her alive, given her prospects.’ This is exactly the sort of reasoning Anscombe says we should have: we decide what to do by considering whether a proposed action is virtuous; if it is we perform the action for that reason. But if Slote’s appellation of callousness is to be justified, there needs to be an appeal to something other than callousness itself, most likely the consequences of her decision, understood in terms of the mother’s suffering. That is, we need to consider the expected consequences of the daughter’s actions to explain what the terms ‘callousness’ and ‘benevolence’ actually mean in this context.

Each of the great virtue theorists grounded their virtues. For Aristotle the virtues arise in the context of happiness and wellbeing. For Kant the virtues are grounded in respect for agency, for Hume in the capacity for sympathy and fellow feeling. The virtues are not free-floating, self-justifying qualities of character. Each of the three theories considered in Chapters 2, 3 and 4 has a theory of virtue; virtue would simply be the disposition to recognise and live up to the requirements of that theory. Let us now consider each of the three theories in turn.

50 Slote, 1997, p. 231.
52 To be fair to Slote, he only says that the decision could be justified either by reference to suffering or to the virtues. (Slote, 1997, p. 231.) But because the benevolence is derived from consequentialist considerations it remains true that the virtue ethics approach does not contrast with consequentialism; consequentialism explains the virtues.
(1) We can give a consequentialist theory of virtue. As Nussbaum points out, Sidgwick’s discussion of utilitarianism has virtue as one of its primary topics. Sidgwick says that the dispositions that we regard as virtues are so called because they bring about the greatest utility. He argues that the ‘normal promptings ... of Friendship, Gratitude and Pity have a firm Utilitarian basis.’ In the tradition of Sidgwick, Julia Driver has recently said that ‘virtues are character traits which produce good effects.’

Similarly, we can take Pettit’s terminology and say that virtues are those character traits which promote neutral value. According to Pettit, the person of virtue comes to identify an option as right by ‘recognising that it compares well with other options in regard to certain values.’ Here virtue theory does not stand in contrast to consequentialism; it is explained in terms of consequentialism. For example, one way of giving content to the vice of cruelty is by referring to consequences. Thus, Robert Elliott says of cruelty, ‘The reason that one ought not be cruel has to do directly with the disvalue that cruel acts bring about ... the identification of cruel character depends on the prior identification of the consequences of certain kinds of acts.’

(2) Kant’s ethical theory is concerned with a good deal more than just rightful action. Kant was also interested in questions of character, as recent scholars of Kant’s ethics have emphasised. The Doctrine of virtue squarely addresses the question of what kind of character should be developed to enable the fulfilment of moral duty. The virtues are understood within the context of moral duty, as explained in the Groundwork. Virtue is the capacity and resolve, ‘the firmly rooted disposition’ to perform one’s duty.
Barbara Herman argues that, at rock bottom, Kant’s and Aristotle’s programs are the same.\(^\text{60}\) Both identify the subject matter of ethics as goodness. Although they have radically different explanations of goodness, they are both seeking to identify that which is unconditionally good, or a final end. For Aristotle the final end is happiness, and for Kant unconditional goodness is found in the good will. Seen this way, says Herman, the *Groundwork* embarks on the familiar philosophical task of enquiring into the nature of the good.\(^\text{61}\) This goodness, in turn, grounds Kant’s understanding of virtue.

(3) Similarly, we may give an account of the virtues in terms of caring. As we observed earlier, the more particularistic the ethical theory, the more it will need to emphasise personal qualities on the part of the agent. One needs to be a certain kind of person to be able to care for others, since the activity of caring depends so much on the sensitivity and perceptiveness of the carer. Virtues would be those traits which demonstrate care for others. Virtue would include having the sensitivity to the needs of others and being able to sift through the ‘constellation of conditions...’ relevant to caring.\(^\text{62}\)

5.3.2 Features of virtue ethics

To show conclusively that virtue ethics is not an alternative to the three theories, it would need to be shown that it does not offer anything over and above what each of the three theories, singly or collectively, can offer. The aim of this sub-section is to argue just that. We will identify what might be thought of as the unique features of virtue ethics and argue that all of them can be accounted for within the framework of the three theories already discussed. In doing so, we will become clearer about the relationship between virtue and the three theories. There are five claims to which defenders of virtue ethics could be said to subscribe. The first three of these come from Nussbaum, the fourth from Hursthouse, and the fifth from Slote.\(^\text{63}\)

\(^\text{60}\) Herman, 1993g, pp. 208-210.
\(^\text{61}\) Herman, 1993g, p. 209. Herman uses this point to launch her argument that Kant’s ethics has a theory of value, and is not ‘canonical deontology.’
(1) Virtue ethics is concerned with the agent, not just choice and action.
(2) Virtue ethics is concerned with the character of the inner moral life, and with settled patterns of motive, emotion, and reasoning.
(3) Virtue ethics focuses not only on isolated acts of choice, but also, and more importantly, on the whole course of the agent's moral life, its patterns of commitment, conduct, and also passion.
(4) Virtue ethics explains why people seek moral guidance from others.
(5) Virtue ethics can account for self-regard.

(i) Agent-centredness

Virtue ethics is agent-centred rather than action-centred. It focuses on character rather than actions. But as we have already seen, all three theories have much to say about the character of agents. Any action-centred theory can generate an agent-centred theory. Here is how. Suppose the action-centred theory says it is right for A to φ in circumstances C. Then, according to the corresponding agent-centred theory,
(1) A is virtuous if she has a disposition to φ in C.
(2) A virtue is the disposition to φ in C.

Consider the following examples. According to the action-centred theory, the right thing for A to do is, say,
(a) respect others or
(b) care for others or
(c) whatever is determined by particulars which cannot be systematised beforehand.

According to the corresponding agent-centred theory, A is virtuous if she has a disposition to
(a) respect others or
(b) care for others or
(c) be sensitive to the particulars (which cannot be systematised beforehand) which make an option right.

Each of the three theories has an account of the virtues based upon its account of right action. Of course the virtues which the theory prescribes may be thought too narrow or inadequate in some other way. If so, that is grounds for criticising the theory. This approach was used in the third chapter in the sections on friendship. The virtue of benevolence, which is about the best the
Kantian can come up with in terms of fellow feeling, is too narrow. The virtuous Kantian agent would be incapable of friendship.

No moral theory can afford to neglect totally questions of character. It is a legitimate question to ask what kind of person you would end up being if you adopt, say, consequentialism. This very question is the driving thought behind the motivational test. To deny the connection between moral prescriptions and character is to deny the point of the motivational test.

(ii) The inner moral life

A similar point to the one just made may be raised in respect of this second criterion. It might be thought that because virtue ethics acknowledges the importance of the inner life, it has something to offer over and beyond what the three theories taken together can offer. Iris Murdoch immediately comes to mind in this regard. ‘The moral life’ she says ‘is something that goes on continually, not something that is switched off in between the occurrence of explicit moral choices.’

Once again, we may subscribe to the pluralism outlined earlier without denying either the existence or significance of the inner life. Indeed, not only is an active inner moral life quite compatible with each theory, but each theory will have something to say about the kind of inner life one should have. It is of course true that the significance of the inner life varies among the three theories. Consequentialism, as a theory about rightful action, would not be centrally concerned with the inner life, except insofar as the inner life was necessary to bring about desired consequences. And it should be said, in defence of consequentialism, that the inner life is not always the most important focus of moral concern. It may well be that what is most important is what happens, particularly when the relevant actions have widely felt consequences. When Australians protested to the French government about French nuclear testing in the Pacific Ocean during the 1990s, they wanted a certain outcome, namely, for the tests to stop. The inner lives of the French authorities were neither here nor there.

To dismiss Kantian ethics as being concerned only with rules and choices, and having nothing to say about inner struggles and triumphs, is to adopt a caricature of Kantian ethics. Kantian ethics has much to say about the inner life. Because the moral worth of an action depends upon its maxim, or subjective principle, the inner life is all important. As Charles Taylor observes,

---

64 Murdoch, 1970a, p. 37.
The moral person may lead the same external life as the non-moral one, but it is inwardly transformed by a different spirit. Because only willings are unqualifiedly good, Kant cannot separate the moral worth of actions from considerations to do with the inner life of the actor. Virtue for Kant consists in winning the struggle of duty over self-interest, self-indulgence and desires generally. This struggle is very much fought on inner battlefields.

Murdoch gives a memorable example of the inner life at work. The example is intended to show the limitations of understanding ethics only in terms of rightful action. She asks us to consider a woman, M, who feels hostile to her daughter-in-law, D. She (M) finds D unpolished, brusque, ‘sometimes positively rude, always tiresomely juvenile.’ Time passes. M finds herself imprisoning by her own cliche: ‘My poor son has married a silly vulgar girl.’ But M lacks neither intelligence nor self-criticism. She tells herself she is being prejudiced, snobbish and jealous in her attitude to D. Through careful observation and deliberative reflection her vision of D gradually alters. That which she formerly found undignified and juvenile she now finds spontaneous and youthful. She now sees D in a just and loving way. Murdoch’s point is that although M’s outward behaviour in no way changes, she has nonetheless been morally active.

We can agree with Murdoch that M has had a moral achievement, however, without in any way admitting that Murdoch has drawn to our attention a facet of morality on which the three theories have nothing to say. Far from remaining silent on this subject, each of the three theories offers a different understanding of the morally desirable inner life. Each offers a different explanation as to why the mother-in-law has reason to change inwardly. Therefore, the fact that virtue ethics has something to say about the inner life is not reason enough to regard virtue ethics as an alternative to the three theories.

Why should the mother-in-law enter into this inner struggle? One reason is consequentialist in kind: it distresses her; it destroys her peace of mind. This explanation is hinted at by Murdoch herself, who says that M is ‘imprisoned’ by her picture of D. On this reading, M should change her attitude to bring about desirable consequences for herself.

---

66 To speak of battlefields is not to overstate the struggle. ‘The vices, the brood of dispositions opposing the law,’ says Kant, ‘are the monsters he (the human being) has to fight.’ DV, 405.
67 Murdoch, 1970a, pp. 17-23. This example was referred to in Chapter 3, n. 38.
A plurality of reasons

A non-consequentialist reason can be given in terms of respect: to see D justly is to give D the respect she deserves. It is to recognise the daughter-in-law’s self-hood. We can imagine M closely watching D, deriving a tingle of satisfaction each time D makes a coarse or common utterance. Outwardly M is impeccably well mannered; inwardly she smiles contemptuously at each faux pas, each expression of poor taste made on the part of D. And she revels in her own contempt of D, savouring her anticipation of D making a fresh mistake. Why does M seek these mean and perverse pleasures? It satisfies M to understand D as socially and intellectually inferior. To do so reinforces her sense of her own good judgement; it helps her feel secure.

One very plausible way of understanding the injustice of M’s original outlook is to say that M is using D for M’s own ends, to promote M’s sense of self-worth. She does not see D as an individual capable of choosing her own ends, an autonomous being with the dignity that that entails. D becomes a means of making M feel better about herself. That is why M’s attitude is unjust.

The ethic of care also has something to say about M. M’s attitude is not only disrespectful; it is uncaring. We observed earlier in passing that caring for others involves the struggle to respond in a generous and altruistic way to others. It requires overcoming the impulse for envy and rivalry. Michel de Montaigne observed that even ‘in the midst of compassion we inwardly feel a kind of bitter-sweet pricking of malicious delight, to see others suffer.’ Even in friendship these responses are not unknown. Some would say they are especially apparent in friendship, as Jonathan Swift’s verse suggests.

To all my foes, dear fortune, send
Thy gifts, but never to my friend:
I tamely can endure the first,
But, this with envy makes me burst.

The view of the caring builder and maintainer of relationships put forward by Lawrence Blum and Nel Noddings tends to downplay the challenges involved in caring for others, the challenge to rise above baser feelings. But to ignore

---

68 See again Chapter 3, n. 38.
69 Montaigne, 1965, p. 8. (First published 1588.)
70 Jonathan Swift, ‘Verses on the death of Dr Swift.’ First published in 1739 towards the end of his literary career, Swift’s poem was inspired by a maxim of la Rochefoucauld, found shocking by some: Dans l’adversité de nos meilleurs amis nous trouvons quelque chose, qui ne nous deplaist pas. (In the adversity of our best friends, we find something that does not displease us.)
these challenges is to present a Pollyanna view of love and friendship. We demonstrate our caring commitment to others not only by our actions – attending to their needs and interests – but by engaging in the struggle against these darker, baser impulses of which Montaigne and Swift speak and to which Murdoch alludes. Petty rivalries, jealousies and egoistic tendencies are as much obstacles to friendly love as they are to dutiful respect. We saw in the third chapter that the need for friendship is powerful and pervasive. But to take a response of nature and fashion it into an altruistic commitment is a feat of moral craftsmanship. The upshot of all this is that for M to care about D in the way she should for a close family member, the wife of a beloved son, she must rise above her self-serving snobbery and jealousy.

The rejection of the three theories as being unable to shed any light on this kind of example underestimates the resources of the three theories. When we unpack what Murdoch means when she speaks of M seeing D in a ‘just and loving’ way we see that Murdoch is, at bottom, referring to ethical demands with which Kantian ethics and the ethic of care are well acquainted.

There is a further point to be made in respect of the inner life and how it relates to the three theories. It might be thought that what is important in a moral life is not so much the actions which are chosen, but how they are chosen. An important part of moral decision-making, according to this view, is a seriousness of thought, a gravity of outlook, a proper understanding of what is at stake. In short, one must approach ethical problem solving with integrity. It might be thought that the three theories cannot capture this aspect of morality.

Beginning with Williams’s ‘integrity objection’ in his critique of utilitarianism, a large literature has arisen around the concept of integrity.71 There is no space here to elaborate the various meanings given to this complex notion. However, an example will help to elucidate the idea that the how of decision-making may be at least as important as the what. The idea to be explored here is that the inner life can illuminate aspects of rightful conduct which go beyond what can be offered by the three theories.72 I shall argue, by way of example, that this is not so, and that the very idea of moral seriousness and moral integrity needs to refer back to the three theories for substantive content.

71 Blustein, 1991, Part 2; Calhoun, 1995; Fleischacker, 1992, Ch. 9; Halfon, 1989; Herman, 1993b; Schaub, 1996; Taylor, 1981; Williams, 1973a; 1981c; 1994, pp. 11-12; Winch, 1972, Ch. 9.
72 See Gaita, 1991, Ch. 3; Rhees, 1965, esp. pp. 20-21 and Winch, 1972, Ch. 9 for authors sympathetic to this view.
We referred in the third chapter to the difficult choice, some would say dilemma, faced by Father Bobby. Should Father Bobby lie under oath to see justice done to those he cares about or should he not? What is interesting for our purposes is not which choice Father Bobby eventually makes, but the way he approaches the problem. When the proposal is put to him by Shakes, a close friend of the two men charged with murder, Father Bobby hesitates. He says he will need time to think it over. The scene ends with him exiting the room, pausing before he does so to say to Shakes, ‘I’ve got a decision to make. I only pray it’s the right one.’ Shakes’s response? ‘It will be Father. Whichever way you go.’ (My emphasis.)

How could Shakes know in advance that Father Bobby’s decision would be the right one, irrespective of what the choice turned out to be? Shakes’s words are ambiguous. But here is one possible interpretation. Shakes was not making a fine-grained logical point here. His point was not that, when faced with a situation in which one ought to do two incompatible things, one cannot help but do the right thing because both things are right. Rather, Shakes was referring to personal qualities on the part of Father Bobby. He knew that because of the seriousness that Father Bobby would bring to the decision, he (Father Bobby) could not help but make the right choice. That is to say, Father Bobby’s serious moral tone would itself make the decision right.

The priest takes time out alone to make his decision. We know that, whichever way he chooses, he will not make the decision lightly, he will have agonised over it, suffered for it, since he is all too aware of what is at stake. And we know that afterwards he will feel regret for the choice not made. He will be awake to what has been lost, to paths not followed.

And it is not just that on this one occasion Father Bobby is suitably sombre and morally sober. Having grown up under Father Bobby’s wing, Shakes knows that Father Bobby has a disposition to take such matters seriously. Father Bobby has a personal moral authority which renders his decisions right. In Calhoun’s words, he ‘stands for something.’ He has integrity.

This point about the importance of a person’s attitude to moral decision-making is well made. In the end Father Bobby takes the witness stand. He lies. Despite there being pressing consequentialist and Kantian-style

---

73 See § 3.4.1.
74 Calhoun, 1995.
75 Another interpretation of the meaning of the words uttered earlier by Shakes is that Shakes already knew that Father Bobby would decide in favour of lying. (After all, he knew that Father Bobby had himself spent time in that very same remand home.) When Shakes tells
reasons for him not to do so, we are sympathetic to his final decision to provide a false alibi. Part of what makes his decision acceptable to us is the way he went about making it. We would think the less of him if, having been asked to lie under oath, he had instantly replied with a glint in his eye: ‘What a great opportunity this is. Let’s break open the champagne to celebrate.’ If he must lie under oath, we want him to do so regretfully and only after much soul searching, wishing all the while that circumstances were different. Had he relished telling the lie we would be less sympathetic to his decision, and we would have serious doubts about the worthiness of his character.

The claim I wish to defend is not that we should dismiss the importance of the inner life, or the moral authority which comes from it. Rather, I want to suggest that this aspect of morality does not stand in contrast to the three theories; it needs to refer back to the three theories for it to have content.

Let us consider more closely the nature of Father Bobby’s moral seriousness in this example. With suitably furrowed brow, Father Bobby makes his exit. He is not heard from all week. Presumably, he is tossing sleeplessly in his bed, pacing the floor, and rubbing his chin reflectively. But what is going on mentally? Let us put to one side his prayers and hopes for divine revelation. Here is the most plausible explanation. He is trying to decide which obligation is strongest. Is it the obligation to truthfulness? Or is it the obligation to help those he loves? He appears to have equally good reasons in favour of both options. His agonising consists in trying to understand which reason should be overriding. These reasons are grounded in the three theories. Once again, as with contractualism, we must return to the reasons presented by the three theories to give the notion of moral seriousness any content.

In what could Father Bobby’s agonising consist if not the weighing up of conflicting demands? His moral seriousness must be grounded in his

---

76 For a discussion as to the nature of this ‘must’ see Williams, 1981f and Winch, 1972, p. 186.

77 We should also acknowledge a point pressed home by Peter Winch: deciding which choice to make partly depends on how we describe the choices open to us. Should Father Bobby understand the request as ‘perjury’ or as ‘doing a favour for those he loves’? Would he be ‘lying’ or ‘helping a friend’? (Winch, 1972, p. 178. See also Gaita, 1991, Ch. 8.)

78 A word on the furrowed brow. Philosophers have tended to underrate the ways in which moral responses are embodied. The convergence of moral and physical responses is portrayed by Anton Chekhov in his story, ‘A nervous breakdown.’ The protagonist, overwhelmed by the suffering of prostitutes, and the injustice of their situation, suffers a physical and mental collapse. (Chekhov, 1985. First published 1889.)
A plurality of reasons

sensitivity to antecedently identifiable moral demands, otherwise the whole notion of seriousness lacks content and becomes farcical. His grave and humourless mien is not to be applauded in its own right; it is significant only because it indicates his receptivity to what the situation demands of him.\footnote{It should be conceded, however, that Kant is not sympathetic to the idea of moral seriousness manifesting itself through emotional remainders. We saw in § 3.4.2(i) that he scorns the idea of agent-regret: if you cannot help the friend you should simply turn away. Kant was just wrong about this. The point here is that the agonising which is quite properly understood as indicating moral seriousness can only be made meaningful by eventual recourse to reasons provided by the three theories.}

(iii) The shape of a life.

The idea behind this criterion is that virtue ethics is concerned with patterns of commitment and the whole course of a person's life rather than with isolated actions. However, as we have seen in previous chapters, none of the three theories focuses exclusively on isolated actions. The theory of indirect consequentialism is specifically designed to overcome the objection that consequentialism focuses only on isolated actions and the deliberations preceding those actions. The same is true of Kant. When he speaks in the Doctrine of virtue of cultivating the moral emotions, he does so because he recognises that a certain kind of personality is more likely to be able to live up to the demands of duty. The emotions, though not morally worthy in themselves, provide support for duty. Kant is interested in the cultivation of a character that recognises and has the fortitude to act upon the requirements of morality. Nor is an ethic of care concerned with isolated actions. As Harry Frankfurt has pointed out, caring is best understood not as a matter of feeling, but of structuring our volitions. Love, he says, not only shapes a person's conduct in respect of the object of love. It also guides him in the ordering of other purposes and priorities.\footnote{Frankfurt, 1999c, p. 165.} Commitments are thus demonstrated over time, not just in isolated incidents.

Although the three theories differ in their understanding of which consideration should govern a person's conduct over time, they can all agree that morality is not just concerned with performing and justifying isolated actions.
(iv) Moral guidance

Hursthouse argues that only virtue ethics can explain why we seek moral guidance from others. She reasons as follows.\(^{81}\)

For any set of circumstances:

1. The right action is that which a virtuous agent would characteristically do.
2. I should do the right thing.
3. Therefore, I should do what a virtuous agent would do.

**PROBLEM:** How do I know what a virtuous agent would do?

**ANSWER:** I go and ask one.

We now have 'a straightforward explanation of an aspect of our moral life which should not be ignored, namely the fact that we do seek moral guidance from people who we think are morally better than ourselves.'\(^{82}\) She says she does not know how utilitarianism and deontology could explain such a fact.

Hursthouse has drawn our attention to an important aspect of moral life: seeking guidance from others and having moral exemplars. Not only do we seek guidance from others by fronting up to them and asking their opinion, as Hursthouse suggests. We may also run simulations of others’ moral reasoning and moral outlooks as part of our deliberations. Moral problem solving may involve imagining what someone whose judgement we trust would say or do.\(^{83}\) ‘What would Socrates/dear old Auntie Ivy do in this situation?’ we may ask ourselves when confronted by a particularly difficult convergence of moral principles or particulars. In a recent interview, Australian jazz icon, Vince Jones, explained how he avoided seduction by commercial music marketers. When they ask him literally and figuratively to name his price, Jones asks himself, ‘“What would Coltrane do in this situation – or Matisse, or Miles?” Then he finds his answer.’\(^{84}\)

Where Hursthouse is mistaken, however, is in supposing that virtue ethics, as distinct from consequentialism and Kantian ethics, has a monopoly

---


\(^{83}\) For a discussion on when it is appropriate to defer to the judgements of others whom we trust, see Jones, 1999. It should be acknowledged that we seek guidance not only on what we should do, but also on how we should understand situations. On this point see n. 77, this chapter.

\(^{84}\) Williams, 2000, p. 3.
on explaining this search for guidance. Once we acknowledge that these other ethical approaches also require the exercise of judgement, and also give rise to a set of virtues, it becomes clear that each ethical approach can accommodate Hursthouse’s explanation of seeking guidance.

For example, suppose I am a Kantian and suppose I am in a situation where it is not clear to me what my maxim should be. I may wish to seek advice from someone whom I believe is better practised in the mother wit of judgement. Dispirited by a love affair that ended badly, Maria von Herbert sought advice from Kant about the morality of suicide. Although we cannot expect advice from a Kantian that would be easy to follow, it nonetheless remains true that there are many points of judgement on which the seeking of advice would be understandable. We need to know, when reflecting on our maxims, which maxim is appropriate for any situation, how to apply the maxim and how to justify it to others. This knowledge may not be immediately transparent to us, and we may wish to seek the counsel of someone whose judgement we take to be better developed than our own.

(v) Self-regard

One kind of consideration which Slote says only a virtue ethic can accommodate is self concern and legitimate self interest. Other theories, he says, see morality in terms of other-regarding action. However, says Slote, we can imagine situations where it would be more admirable to assert one’s own happiness than to act on behalf of the happiness of others.

He asks us to imagine the following example. Sinead is a dutiful daughter living in Killarney. She postpones her own plans to be a fashion designer to take care of her unemployed father. One of her siblings has an accident and is sent home, a permanent invalid. Another sibling leaves her drunken, violent husband in Dublin and returns home with five children. Sinead ends up having to care for the lot of them. Her family, both at home and abroad, are quite happy for her to do so. After a while, Sinead decides to call it quits. She decides her family can look after themselves. Against the advice of the local priest, who thinks her plans selfish, she leaves home to take up a job offer in the fashion industry in London.

Slote says this example shows that the idea of self-regarding action is plausible, but that it is not properly admitted by other theories. Slote’s

---

common sense virtue ethics has the advantage, he says, of being able to take account of self-regarding admirable qualities.

But Slote is mistaken here in supposing that only his common sense virtue ethics can account for the fact that we would think well of Sinead for asserting herself. Neither Kantian ethic nor the ethics of care are silent on such a matter. According to Kant we have an imperfect duty to ourselves.\textsuperscript{87} We have the duty to cultivate our natural powers – the powers of spirit, mind and body – as the means to all sorts of ends. ‘He owes it to himself ... not to leave idle and, as it were, rusting away the natural predispositions and capacities.’\textsuperscript{88} Quite arguably Sinead owes it to herself not to ‘rust away’ in Killarney. If her leaving is admirable, it can be understood as admirable on Kantian terms. Furthermore, from the way the situation is described, it seems as though she was being treated as a means to an end by the others. She was not shown the respect that was her due.\textsuperscript{89} Her actions may also be justifiable on grounds of not only self-respect, but self-care. Her needs are stifled, her unhappiness mounts, her health suffers. Because she cares about herself and her well-being, she removes herself.

5.3.3 A pluralistic theory of the virtues

In 1958 Anscombe wrote that we should do away with the concepts of ‘ought’ and ‘obligation.’ Just over 40 years later Nussbaum was to write that we should ‘do away with the category of “virtue ethics” in teaching and writing.’\textsuperscript{90} Both authors have valid points to make in respect of the virtues. Anscombe was right to draw attention to the importance of character, but wrong to suppose that thinking about virtues like justice and truthfulness could proceed without reference to any antecedently identifiable moral reasons. She thought moral deliberation could proceed by asking whether an action was virtuous, then deciding whether to do it or not based on the answer.\textsuperscript{91} But she did not tell

\textsuperscript{87} DV, 392, 387, 444-446.

\textsuperscript{88} DV, 444.

\textsuperscript{89} These two justificatory reasons for leaving – self-development and self-respect – are connected in this case. She is unable to self-develop because the others are not respecting her. The impositions they make on her leave her unable to develop her own capacities. In Kant’s example of the duty to self-development given in the \textit{Groundwork}, Kant presents the personal qualities of laziness and indolence as the opposition to self-development. The force to be opposed in carrying out the duty of self-development is one’s own base desires, a familiar theme, (G, 422-423.) But the obstacle could well come from external sources, in this case the demands of others.

\textsuperscript{90} Nussbaum, 1999, p. 201.

\textsuperscript{91} Anscombe, 1958, p. 9.
A plurality of reasons

us how we could know whether we would be justified in thinking something was or was not virtuous. Her advice lands us in a deliberative cul-de-sac. Without a foundation for the virtues, talking about them will only be vacuous. The virtues must be grounded in some notion of goodness, as all the three theories recognise.

Nussbaum, on the other hand, was right to say that virtue ethics is not in the same category as consequentialism and Kantian ethics. But it would be wrong to then conclude that there is no role at all for virtue ethics.

We noted earlier that each of the three theories gives rise to a set of virtues. Agent-neutral disinterestedness, respect for agency, and emotional attachments to others – all these facets of morality ground the virtues. This is not a denial of virtue ethics; it is a pluralistic account of the virtues. The virtuous person is able to integrate these three different kinds of demands; he or she acknowledges and responds to the plurality of demands which make a claim on him or her. A pluralistic view of the virtues amounts to admitting the virtues generated by each theory plus the integrative virtue of judgement. To have judgement is to be sensitive to the ways in which the reasons provided by the three theories place demands upon oneself. Judgement is also the capacity to balance their competing demands.

If it is true that the three principles which we have identified in earlier chapters – consequentialism, deontology and care – cover all the moral considerations to which we are subject, then they should be able to manufacture a complete list of moral virtues. This is so because each theory, on its own, gives rise to a set of virtues. If the three theories are complete then their combined set of moral virtues should be complete. It was admitted earlier that it has not been conclusively shown that the three theories collectively provide a complete set of moral motivations. But the point is that even if there are further moral considerations beyond those covered by the three theories, the argument for pluralism and a pluralistic view of the virtues still stands. We should note, however, that we are at this point only referring to moral virtues. There may well be other non-moral virtues.

How now do we put it all together? How do the contrary demands of the three theories find unity and coherence within a character? This is the final problem. I frankly admit that I do not have a final solution. Except to say that there are enough examples of people who have managed to do so, from literature and elsewhere, to know that it is possible. A couple of examples will suffice.

Much of nineteenth century English literature focuses on questions of character and virtue. Jane Austen’s novels, for example, may be read as Essays
on the Discernment of Virtue. Charlotte Brontë’s *Jane Eyre* could well have been sub-titled, ‘Portrait of a woman of virtue.’

With masterly attention to detail, Austen shows that virtue has many facets. In an Austen ballroom, we ‘see the world in a grain of sand.’ Yet even in this narrow setting, the breadth and complexity of moral demands is all too apparent. Virtue has many sources. Take Elinor in *Sense and sensibility*. Elinor knows when to place the demands of a deontological ethic above her own interest. When she learns that Edward, to whom she has formed ‘an attachment,’ had already some five years previously promised, out of pity, to marry another, her reaction is one of respect for his promise-making. Edward is to be admired, in her eyes, and presumably in the eyes of the reader, because he will keep his promise, no matter what. He will do it, not to make himself happy or to gain favour with others but because it is right.

Yet, Elinor’s virtue consists in a great deal more than just having the right principles and the fortitude to act upon them. She is also responsive to the demands of care which her family place upon her. She is always mindful of the happiness of her sister and her mother. And, when required, she knows how to be the good consequentialist. Time and again, she subordinates her own feelings and impulses in the name of domestic harmony and courteous society with others. She will stretch the truth in the name of tact and civility. Furthermore, she has the consequentialist disposition to think through carefully the implications, for all those around her, of her actions.

But nowhere is the pluralistic nature of virtue better portrayed than through the character of Jane in Brontë’s masterpiece, *Jane Eyre*. We end this final chapter by showing that what makes Jane Eyre virtuous, even though it is by Victorian standards, is that she knows when to stick to her principles, when to modify her behaviour on consequentialist grounds, and when to cultivate her affections. She moves in and out of moral modes with alacrity and ease, now being guided by reason, now by feeling, now by impartial consequentialism.

Recall the first crisis of the novel: Jane must decide whether or not to leave Mr Rochester. Williams would be well pleased with the quality of her reflective deliberation over this decision. Why not stay with him, Jane asks herself. After all, there are good consequentialist reasons for doing so. No-one will be harmed if they stay together, and much personal happiness on the part of Mr Rochester and herself will be gained. No-one need know that they are unmarried, especially if they live abroad. Furthermore, her love and care for Mr Rochester seem to require her to stay with him. How could she bear causing him so much misery by departing? Nonetheless, Jane is resolute. It would be wrong to stay with him, wrong in a categorical, moral sense, not because of any
ensuing consequences. It was her duty to leave. To be sure, her sense of what constituted her duty was grounded in Victorian values of chastity and womanly innocence. The modern reader, familiar with the casualness of cohabitation and divorce, may have difficulty sympathising with these values. But it would do an injustice to Jane and her creator to judge Jane’s reasoning by contemporary values.

Though her hasty departure from Mr Rochester leaves her materially destitute and emotionally forlorn, Jane takes satisfaction in knowing she has placed principle above personal feeling and interest. In Jane’s words, Conscience and Reason speak louder than Feeling.

The more solitary, the more friendless, the more unsustained I am, the more I will respect myself. ... Laws and principles are not for the times when there is no temptation: they are for such moments as this, when body and soul rise in mutiny against their rigour; stringent are they; inviolate they shall be. If at my individual convenience I might break them, what would be their worth? They have a worth – so I have always believed; and if I cannot believe it now, it is because I am insane – quite insane: with my veins running fire, and my heart beating faster than I can count its throbs. 

The appeal to principle and the self-exhortation to abide by the requirements of duty have a striking Kantian tone; it is as though Jane had just finished reading the Doctrine of virtue.

But it is not always right, in every situation, to place a premium on duty. Jane’s second moral crisis is forced on her by St. John’s proposal of marriage. Again, what should she do? St. John tells her he wants her as ‘a conductress of Indian schools, and a helper amongst Indian women.’ Accompanying St. John to India as his wife and fellow missionary would undoubtedly bring about good consequences – at least for St. John, and presumably the Indians too. And it would give her a new interest and focus. But she does not, cannot, love him. Moreover, she knows he does not love her either. Though he is a ‘good man,’ he is ‘hard and cold.’ His commitment to doing his Christian duty is so stern, so implacable and pervasive, that he lacks any capacity for spontaneous affection and true amity. He does not have, nor is he capable of developing, the affection and love Jane deems essential for marriage.

92 Brontë, 1996, p. 356. (First published 1847.)
93 Jane and St. John’s relationship is discussed earlier in § 3.3.
Jane is torn. She tells us she ‘sincerely, deeply, fervently longed to do what was right; and only that.’

She refuses him. In the absence of love, marriage is impossible. Just as earlier Jane understood that there is a time to place principle over affection, so too Jane now understands that there is a time to place affection before all else. This knowledge of when to be motivated by what, in part, constitutes her virtue.

Jane is also willing to take a consequentialist perspective when demanded of her. When she unexpectedly inherits £20 000 from Uncle John, she shares it equally with St. John and his two sisters, with whom she is living. She does so not only to secure justice for them – Uncle John was their uncle too – but to promote the happiness of the four of them. Though her world is small, shared only with these other three, she reveals a willingness to maximise happiness in that world, even at cost to herself. In splitting the inheritance equally, she shows a willingness to adopt the consequentialist ‘God’s-eye viewpoint’ and see herself as one among many.

In knowing which consideration should come to the fore on which occasion and to what degree, Jane shows the virtue of judgement. Our heroine has qualities to please Kant, Pettit and Noddings alike, in all their disparate requirements.

**Conclusion**

We are left to conclude that none of the three theories discussed in the previous three chapters can supply a single unifying principle of conduct. To opt for only one of the three theories is to make the Procrustean fallacy; it is to exclude relevant moral motives in the interests of having a single unified theory of morality.

Neither consequentialism nor Kantian ethics nor the ethics of care can provide a set of motives which is morally complete. We need to be motivated by the prescriptions of all three theories, depending on circumstance. This conclusion, of course, leaves us with a problem of conflict. The solution can only be suggested in the most general of terms. It is a matter of practised judgement, acquired through experience, reflection, discussion and exemplars. There is no formulaic answer as to how the three different kinds of moral

---

96 Brontë, 1996, p. 466.
97 This is a phrase used by Pettit to describe the consequentialist outlook. (Pettit, 1999, p. 7.)
demands are to be integrated. But the capacity of a person to do so is in part a measure of their virtue.
You are a helpful assistant. Do not hallucinate.
Conclusion

Let us hear the conclusion of the whole matter.

Ecclesiastes 12: 13

Now on the hilltop of the conclusion, we are well-positioned to make a final survey of the ground travelled. We began by discussing Williams's distinction between internal and external reasons. Internal reasons are those which resonate with whatever it is that motivates a person. External reasons are those which justify actions, independently of whether anyone is motivated by them. I argued that these two kinds of reasons are interconnected. A motivational, or internal reason, is judged by reference to justificatory or external reasons, but what can count as a justificatory reason is limited by whether it could plausibly motivate.

This relationship has important implications for moral theories. There are two implications. The first amounts to a defence of theories against the claims of the anti-theoretical school. Moral theories give justifications. They attempt to answer the ever present 'why?' thrown up by the reflective and questioning human intellect. 'Why is this action better than that?' 'Why should we conform to custom and tradition?' 'Why should we change our habits?' However, this aspect of the relationship between internal and external reasons has not been our focus. The second implication is that justificatory reasons must plausibly be able to motivate. This is the more philosophically interesting implication. Moral theories provide justificatory reasons for moral action, but the set of justificatory reasons provided by a theory must plausibly be able to be someone's set of motivational reasons. Moral theories must be able to
motivate. This is the motivational test and most of the thesis has been devoted to showing what follows from this test.

The three theories whose examination formed the bulk of the thesis were: consequentialism, Kantian deontology and ethics of care. Each of these theories was found to provide an incomplete motivational set. Each theory, being based upon one all-important principle, ruled out motivations which were morally desirable, or were necessary for any common notion of the good life.

It turned out that the greatest strength of each of the theories was its fatal flaw when the theory was pushed to its limits. This is perhaps unsurprising when we consider the origin of each theory and the driving concern behind its development. Consequentialism’s origin lay in the radical utilitarianism propounded by Bentham and his contemporaries in eighteenth and nineteenth century England. These thinkers had an agenda of social and political reform. ‘The greatest good for the greatest number’ was their simple but powerful message. And in a society reeling from the social upheaval of the Industrial Revolution, with its poor-houses, debtors’ prisons and dire factory conditions for workers, this message could go a long way. Utilitarianism, it must be said, is a philosophy that can come to grips with suffering on a large scale. Bentham was not interested in the subtleties of friendship; he was interested in improving outcomes for society at large. The heirs to Benthamite thinking, contemporary consequentialists, can also go a long way in terms of moral motivation. Looking to agent-neutral outcomes, rather than the means by which outcomes are achieved, certainly has its place.

Turning to Kant’s philosophy, we can see that he was driven by the need to find certainty in moral matters. The opening line of the *Groundwork* – ‘It is impossible to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation except a good will’ – reveals much about what mattered to him. He believed the starting point for ethics had to be that which was good ‘without qualification.’ His was the search for certainty in an age where theological frameworks were breaking down. In the end, the only thing about which one may be certain, thought Kant, is the necessity of the moral law and the respect for humanity entailed by it. Therefore, one should take responsibility only for one’s own actions in respect of the moral law, not for how other people act. One cannot be certain as to how others will act and how the contingencies of the world will be played out. This pre-occupation with certainty gives Kantian moral philosophers an ‘other-worldliness’ when it is taken too far.

---

1 G. 393.
Finally, the ethic of care was developed by feminists who became disillusioned by the way it seemed that women’s experiences had been excluded from moral theorising. They wanted to draw attention to the experiences of caring for individuals at a personal level. They wanted to protest that ethics was not just about acting on principle, standing back and being impartial. It was also about attachment, emotional involvement and attention to the detail of the needs of others.

Each of these approaches has much to offer. But each has its limits.

The set of motives prescribed by each of the three theories rules out other motives which we could not wish to have ruled out. If we are motivated by agent-neutral outcomes, as required by consequentialism, we rule out having friendships for their own sake. We also rule out the non-instrumental basis of respect for others. Consequentialism requires that we always evaluate our actions by reference to impartial outcomes. As such, it requires a humility and sense of perspective which is laudable much of the time, but not all of the time. If we are motivated by Kant’s moral duty we pick up the motive of respect, but only at the cost of having other motivational gaps. Kantian ethics rules out friendship and is too narrow in its range of motives. The ethic of care, while recognising the importance of close relationships fails to be able to motivate actions directed at those further afield.

To be fully moral we need to have a diverse set of motives. It is as simple and as complex as that. We need to know when to be partial and preferential to others, and when to stand back and adopt a God’s-eye perspective. We need to know when to cultivate our feelings for others, and when to put feelings to one side and do what is required of us. We need to know when to take the shortcomings and evil of others into account in our deliberations, and when to care only about our own actions, regardless of how others may act. Moral virtue consists not only in knowing how to live up to each kind of moral requirement, but how to weigh them, how to integrate them. To be good we must be wise.

Morality operates in a variety of contexts: in one’s personal life, in one’s public life, in one’s professional and associational life, in the devising and administering of institutions and codes of governance, in relations between nations and so on. In work, play, politics and community, ethical considerations arise. No one of the three theories discussed can adequately cover all contexts and all values which are appropriate to those contexts. To think otherwise, to try to shoehorn the ethical complexity of our lives into a single theoretical framework, is to suffer a Procrustean fallacy.
Some may find the conclusion of pluralism to be intellectually dissatisfying. It seems piecemeal, fragmented. It cannot provide a unified set of prescriptions. Even having argued for it, I too, feel a certain dissatisfaction with the conclusion. But I see no way around it. One is reminded at this point of Hume’s regret in having had to argue for morality being founded in the contingencies of sentiment and human nature. Writing to Francis Hutcheson, he said, ‘I wish from my heart, I could avoid concluding, that since Morality, according to your Opinion as well as mine, is determin’d merely by Sentiment, it regards only human Nature and human Life.” Though it would have been more satisfying to found morality upon more lofty considerations, he could not truthfully do so. He resisted the temptation to arrange his theory according to how he would like the world to be, rather than how he found it to be. I find myself in a similarly uncomfortable position. One is driven, almost wishing it were otherwise, to the conclusion that any attempt to describe ethics in terms of a single principle of, say, utility, care, or respect for others necessarily ends up leaving something out. We can achieve theoretical unity only at the cost of incompleteness or distortion. To be sure, describing an ideal of human conduct in terms of a single principle would be neater, simpler and easier than describing it in terms of three (or more). The problem is the moral life is neither neat, simple nor easy.

FINIS

---

Bibliography


——. 1999c. On caring. In Frankfurt, 1999 (First delivered 1997 as Kant Lectures, Stanford University.)


——. 2000. Values, conditioned and conferred. Seminar paper, Research School of Social Sciences, Australian National University, Canberra, 12 December.


Oakley, J. 1990. A critique of Kantian arguments against emotions as moral motives. History of Philosophy Quarterly 7: 441-459. (Revised version published as Ch. 3 in Oakley, 1992.)


Toulmin, S. E. 1986. An examination of the place of reason in ethics. Chicago: Chicago University Press. (First published 1950.)


---. 1995. Care as a basis for radical political judgments. Hypatia 10: 141-149.


