Being good with feeling:

*Some problems about the place of emotion in moral agency*

Sophie Rietti

*Thesis submitted for the degree of Ph.D. in Philosophy at the University of Edinburgh*

2002
The enclosed is my own work except where otherwise stated in notes or references.

Sophie RDA
Thesis abstract

The role of emotion in moral agency: Some meta-ethical issues in the moral psychology of emotion

This thesis aims to elucidate an apparent paradox about the role of emotion in moral agency. A number of lines of concern suggest emotion may have serious negative impact on moral agency. On the other hand, there are considerations that suggest emotion also plays a crucial role in motivating, informing and even constituting moral agency. Significantly, there is a strong connection between participant reactive attitudes and ascription of moral status as agent or subject. Non-emotional agents could not hold such attitudes. Also, removing participant reactive attitudes imposes a peculiar and incoherent form of solipsism about moral agency.

Given this necessary role for emotion, can we give an account of emotion that will also meet the worries? I examine, as crucial examples, three recurrent lines of concern about emotion - that it threatens our capacities for objectivity, rationality, and autonomy - to tease out the descriptive assumptions about emotion, and the normative assumptions about moral agency, that these objections are based on. I then offer three lines of argument towards resolving these worries. The first addresses the worries directly, and the other two shift blame off emotion.

First, then, I argue that the normative concerns can largely be met by a descriptive account that views emotion as cognitive. However, “judgementalist” cognitive accounts that assimilate emotion to belief may make emotion meta-ethically respectable at the cost of making it meta-ethically redundant. Also, such accounts are descriptively less than plausible. A better approach, I argue, is to allow that belief may play a significant role in emotion but to also allow at least a quasi-cognitive role to the distinctively affective element in emotion: feeling.

I also argue for a further revision of cognitive accounts to emphasise that emotions reflect features of those who feel them. If we were different, our emotions would be different. So, secondly, I argue that a number of the features that power worries about emotions have their sources in what those who feel them are like, rather than in emotions as such. However, both human nature and emotion are capable of significant plasticity and diversity. We are also capable of a considerable - but not infinite - degree of self-determination both about what we are like and what our emotions are like.

Finally, I argue that the normative assumptions that power the objections to emotion are themselves in need of revision - and in some tension with each other. This leads to a McGuffin-theory of emotion in moral agency: Problems with emotion’s place in moral agency serve as indicators of unresolved tensions in our thinking about moral agency, rather than just indicators of problems with emotion as such. In view of this, I also argue for caution in any attempts to change emotion to fit particular ideals of moral agency.
Acknowledgements

This thesis builds on work done for my first (Master’s) research degree. Thanks are due to Dory Scaltsas for supervision of my Master’s thesis, and also to Andrew Mason, Geoffrey Madell and Michael Menlowe for supervision and support for shorter pieces written for the Master’s degree and which have contributed toward later Ph.D.-thesis work. Professor Scaltsas also supervised the first year of my Ph.D., when this thesis was still a more specifically Aristotelian project. During that year, I also benefited from working jointly with Dr. Madell on emotion in philosophy of mind and with Dr. Jonathan Jacobs, then visiting fellow at Edinburgh, on emotion in meta-ethics and moral psychology. Last but not least, thanks are due to Richard Holton, who took over my supervision from my second Ph.D.-year and has seen this project through drafts, drafts of drafts, and final drafts.

Some of the arguments and ideas that went into this thesis come from other sources than my main Edinburgh work. My Strawsonian arguments in Chapter One go back to my first postgraduate effort: a short paper written for a research seminar run by the Norwegian State Research Council in Oslo, Autumn 1997, headed by Martha Nussbaum and Dr. Arne Johan Vetlesen (University of Oslo). A later version of those arguments, developed through thesis work, was presented at the Royal Institute of Philosophy conference on emotion, Manchester 2001.

The main argument in Chapter Six was presented at a postgraduate conference on values at Southampton in early 2000, as a postgraduate paper at the Joint Session later that year, and for a Work-in-Progress seminar at Edinburgh in early 2001. I am grateful to all those who attended these papers for feedback, comments and advice, and to various conference organisers for forwarding feedback from selection committees and audiences. Special thanks for particular follow-ups to Al Mele, Jonathan Wolff, Jonathan Dancy, Tim Chappell and David McNaughton.

Throughout my postgraduate work, I have benefited from seminars, reading groups and presentations at the Department of Philosophy at Edinburgh University: Thanks go to staff and fellow students for stimulation and feedback throughout. I am particularly grateful that several of these people, many of them with little or no relation to my area of specialisation, have gone out of their way to alert me to relevant readings and events. Some of those who helped somewhere along the line: Ben Young, Richard Gray, Hild Leslie, Alastair Renton, Brian Schaeffer, Emily Postan, Kimberley John, Nikos Kakalis, Stasinos Stavrianeas, Eleni Manolakaki, Julian Kiverstein, Trina Ferguson, and the McDowell reading group.

At Edinburgh, thanks are also due to the Philosophy of Department secretaries, the staff at the Faculty of Arts, and the ArtLAN computer support staff, for student-friendliness.

Thanks also to Grethe, Pete, Grant and Stasinos K., for support in various minor computer-related crises.

And finally, thanks to friends, family, significant others through the postgrad years, for support, advice, diversion, sanity-revival and occasional hindrance.
Introduction

Aims and recurring themes

Recent literature has seen a resurgence of interest in the emotions, both within philosophy of mind and in moral philosophy. This resurgence of interest, moreover, is predominantly positive. In philosophy of mind, emotion is increasingly seen as an interesting and complex cognitive phenomenon, as opposed to some kind of brute boo-hurray phenomenon. In moral philosophy, largely due to the revival of virtue ethics, emotion is now more often seen as a force for good in moral agency and a crucial bearer of human goods.¹

Broadly, this thesis falls within the current tradition, insofar as it defends a place for emotion in moral agency, and does so in large part on the basis of reading emotion as a cognitive phenomenon with important similarities to beliefs and judgements. The starting point here, however, is not the cognitive account as such: rather, the cognitive account plays an explanatory and supporting role to some more basic intuitions about the place of emotion in moral agency. The aim of this thesis is to examine and develop those intuitions.

A number of more general themes recur throughout the thesis. First of all, there is a paradox about emotion and moral agency. While I shall argue that emotion plays a crucial role in moral agency, and in our understanding of moral agency, traditional lines of objection to emotion that view it as a pernicious influence on moral agency also have some descriptive and normative plausibility.

¹ For early arguments against a reductive, non-cognitive reading of emotion in philosophy of mind, see e.g. Bedford (1957), Alston (1967). Emotivism, in its early inception, used a reductive account of emotion to underpin a reductive account of moral judgement (Ayer 1936, Stevenson 1937) - though see Williams (1973b) Urmson (1968) and Wiggins (1992) for revisionist readings of emotivism that make it non-reductive about both emotion and morality and view it as a basis for giving emotion a central role in moral agency. Independent arguments on similar lines can be found in Gibbard 1990 and Blackburn 1998.

Early emotion-positive arguments in moral philosophy can also be found in MacMurray (1935) and Broad (1954). The virtue ethics-associated case for emotion in moral agency is usually traced back to Anscombe (1958) and Foot (1978). For similar approaches outside the dominant Aristotelian/Humean paradigm, see also G. Taylor (1975) and Neu (1978). Recently, of course, there has also been a resurgence of folk interest in emotion focused around popularised psychological
Secondly, although the paradox can largely be resolved, some of the features that make emotion problematic are also ones that make it crucial for moral agency.

Thirdly, some of the features with this double effect seem to be inherent to emotion - but more often, I argue, blame, or credit as the case may be, lies with sources other than emotion itself.

Fourthly, defenses of emotion against one line of attack sometimes land it in further trouble with another traditional line of objection, which leads to the suspicion that emotion is working as a McGuffin for unresolved tensions in our assumptions about moral agency - and in some of our more general meta-ethical assumptions. In other words, emotion functions as a focus for discussion of issues that are not necessarily specific to emotion, or at least are not just about emotion.

A subsidiary point related to this is that our higher-level capacities for self-regulation and self-determination, such as reason, may themselves be flawed and vulnerable to antinomies. It may also be that they, rather than the lower-level phenomena that we use them to control, are in the wrong.

Outline and chapter contents

I start, in Chapter One, by arguing from the implausibility of the extreme view that it might be better for moral agency to remove emotion altogether. Such views have historically real sources, but I will not deal with the full-scale historical issues here. Instead, I will be using a strawman: the (lower-case) stoic, who, not being a historical Stoic, may also draw on anti-emotion arguments of other kinds and later periods than his historical namesakes.

As emotional creatures we are prone to cognitive and motivational distortions, to feeling-based hedonic upheaval, to vulnerability to things outside of our control, and to loss of control from inside. Removing emotion might be thought to remove all these problems. However, the stoic strategy of improving moral

---

work such as Goleman's 1995 best-seller *Emotional intelligence.*

A McGuffin is an entity, like the Maltese Falcon in the book and film of that name, that is the ostensible focus of an enquiry whose real point turns out to be something else. The term is usually attributed to Hitchcock, who often used McGuffs centrally in his films - the entire Janet Leigh
agency by removing emotion, and the capacity for feeling, altogether, is one of throwing the baby out with the bathwater.

Most significantly, this is so because our capacity for feeling emotion underpins what P.F. Strawson calls participant reactive attitudes. And, I shall argue, there is at least conceptually a crucial connection between being a legitimate target of such attitudes and being a moral agent - and between being someone on whose behalf they are held and being entitled to moral consideration. In view of this, I argue that the stoic approach is not only implausible but also incoherent as a way of improving moral agency. That emotion plays a crucial role in moral concerns and agency is also supported by two further arguments: 1) That emotion may play a crucial and irreplaceable distinctive role in moral perception and motivation; 2) That without our capacity for feeling we would only be capable of very attenuated, if any, flourishing.

These arguments offer a global defence of emotion, and leave open the possibility that emotion can still be faulted, not just locally but also globally. Emotion may still be troubling not just where it goes wrong in individual instances, but also in more general terms - which typically cashes out as concerns about its kind-features. Moreover, if we cannot do without emotion in moral agency, then there is surely all the more reason to get clear about what problems this necessity might bring with it. And all the more so if we are to have some hope of seeing how those problems are to be managed.

In view of this, Chapter Two examines three main traditional lines of objection to emotion: From threats to objectivity, rationality and autonomy. Specifically, I will look at four kinds of subjectivity that have been attributed to emotion: Phenomenological, brute feeling, projective and agent-relative, and why they might be thought problematic. On the point of rationality, I will look at how emotion stands with respect to standards of cognitive and strategic rationality, and how these standards themselves feature in morality. I will also look at two different kinds of worries about what Kant calls heteronomy: loss of autonomy through the subplot in Psycho could for instance be viewed as a McGuffin.
impact of externals, and loss of autonomy from within through loss of self-control or through our perceived passivity with respect to our own emotions.

The aim is to tease out both the descriptive assumptions about emotion, and the normative assumptions about moral agency, that power these lines of objection, and to see how they may be met.

The hunch at this point is that while the objections have some degree of plausibility, they are overstated, and can in large part be met. Some of the standards by which emotion is thought to fail are themselves problematic. Also, there are tensions between some of the normative assumptions that power concerns for objectivity and some of those that power concerns for autonomy. And both of these concepts have two-way connections with standards of rationality, which may themselves be subject to some internal tension - for one, cognitive and strategic rationality may not always go together.

In order to have some idea of how far the objections can be met, and also to get some clearer understanding of why and how emotion should play such a crucial role in moral agency, we will need to have some clearer idea of what emotion is. This is the topic of Chapter Three.

I focus discussion around the currently dominant descriptive approach to emotion in philosophy, the so-called cognitive theory. Or rather theories, since there are a number of distinct versions. Such approaches are particularly relevant here as proponents of cognitive theories, insofar as they also have interests in moral philosophy, tend to take it that viewing emotion as a cognitive phenomenon provides a basis for a moral-psychological vindication of emotion.

Cognitive accounts of emotion achieve this, roughly, by arguing for emotion as involving cognitive appraisal of intentional objects: such appraisals are assumed to be responsive to, and even follow from, appeals to relevant explanatory or justificatory reasons. Such reasons are in turn usually assumed to be specific to particular emotion-types. This account would seem to bring emotion within the realm of minimal rationality. Possibly, insofar as emotion-grounds provide standards of intelligibility and appropriateness, this account of emotion will provide some
hope for a degree of objectivity. And insofar as emotion is supposed to be responsive to reasons and standards in this way, this account also gives some hope of meeting worries about loss of autonomy, if only (!) by rendering emotion rational.

Broadly, I think this kind of approach is workable, but with some important qualifications relating to my earlier arguments. For one, I argue that those accounts that assimilate emotions to other cognitive kinds such as judgements and beliefs buy meta-ethical respectability for emotion at the cost of redundancy. For if emotion turns out to be just doubling up work done by some other faculty, such as reason, desire, or both, cognitivist accounts of emotion could hardly make a case for the necessity of emotion in moral agency. If emotion does have a necessary role to play, it will have to be in virtue of some distinctive traits of its own. And, last but not least, strong cognitivist accounts have some serious plausibility-problems on the descriptive side.

These problems are, I argue, mutatis mutandis inherited by more moderate kinds of cognitive account, insofar as these also tend to lend emotion respectability in virtue of its non-distinctive rather than its distinctively emotional traits. Broadly, what is needed to explain, and hopefully also vindicate, a crucial role for emotion in moral agency is a cognitive account, but one that takes on fully the specifically emotional nature of emotion.

Specifically, such an account needs to make room for the possibility that emotion is a sui generis mode of cognitive awareness, whose cognitive content cannot be simply captured in terms of other kinds of cognitive states such as beliefs. I also argue that feeling may be doing an emotion-specific kind of cognitive work. This emphasis on feeling does not preclude an important role for standing or dispositional emotion, of which feeling may be a less immediately prominent (though not altogether dispensable) feature. Nor does it preclude an important role in emotion for other cognitive kinds such as beliefs.

Further, an account that takes feeling as crucial emphasises emotion as a state the subject is in as well as a state that is about an object. And, even independently of that point, I would argue, as Aristotle’s does in The Art of Rhetoric, that emotions reflect typical features of those who feel them as well as
what they are felt towards. If this is so, we may be able to draw inferences about the likely nature (or at least temporary state) of the subject from the emotion, the person that does the feeling. Conversely, what we know about the subjects of emotions helps us predict their emotions. Grounding emotional reaction in the nature of the subject as well as of the object of emotion also helps diminish the apparent randomness of emotional reactions.

Furthermore, this notion of emotion reflecting subjects as well as objects helps to shift some of the responsibility for the features of emotion that power the objections onto sources external to emotion itself. Specifically, I argue in Chapter Four that there are important sources of heteronomy and "bad" subjectivity in human nature and the human situation, and that to a great extent emotion is merely reflecting this. Though it needs to be conceded that emotion, in reflecting such features, may also reinforce and amplify their impact.

Insofar as the specific forms our emotions take will reflect the kind of subjects we are, and our situation, it is worth emphasising some basics about our nature and situation. For one, we are quantitatively and qualitatively distinct and separate from each other, and from our non-personal environment. These quantitative and qualitative distinctions - which metaphysically speaking may be entirely contingent - have fundamental epistemological and phenomenological consequences. Our separateness and distinctness implies idiosyncratic, egocentric and even solipsistic tendencies. And we are, apart from these biases, in more general ways limited and fallible both as perceivers and as reasoners. We are also dependent on our environment and on others (and their attitudes) for survival, let alone flourishing, which makes us vulnerable to externals generally and to the attitudes as well as actions of other agents specifically.

So we have a number of potential sources of bad subjectivity, of rationality-failures, and of heteronomy, built into our nature and situation before emotion is even brought into the picture. And insofar as all these features are present, we would expect them to be reflected in our emotions. But insofar as we are able to imagine their emotional lives at all, creatures whose situation differs from ours in the
relevant respects would have different emotions to ours.

Now, we often deal with the challenges presented by these basic features of our situation more or less badly. And it needs to be conceded that emotion does not always help, and may even sometimes reinforce the problems. But it is not by itself the source of them. And insofar as emotion works with what it is given, we should perhaps look elsewhere than to removal of emotion for ways to sort the problems out. Rather, I suggest, what is needed is some fine-tuning of our emotional and conceptual frameworks (which anyway tend to feed into each other) to deal with the challenges better. And in fact, as humans, we possess capacities for critical reflection and creative self-determination that can also be applied to transforming our emotions. In some of these capacities, such as empathy and sympathy, our capacity for feeling itself plays a crucial part in bringing about the transformation.

Further on these lines, Chapter Five, picking up a thread from Chapter Three, looks more closely at the role that ideologies may play in informing ways of controlling and shaping emotions. I look at what limits there are to our emotional flexibility and hence to the malleability of emotion to ideology. I also look at what limits there might be to the desirability of such changes, and at some of the pitfalls that such attempts at change face. This chapter picks up more explicitly on two threads that have been raised in earlier chapters. 1) The uncertain meta-ethical status of what is given in our nature and situation, and 2) the worries we might have about the flaws in our higher-level capacities for self-determination and self-correction.

Finally, picking up the McGuffin theme again, there may also be underlying tensions in the kinds of normative assumptions about moral agency that motivate objections to emotions. Chapter Six takes a closer look at one way in which these tensions manifest themselves: In the way that differences between agents in their moral characters and judgements raise meta-ethical problems.

Insofar as such variation is thought to be a bad thing, objections are often raised to emotion as a source or constitutive element of it. And insofar as it could be argued, as I do in Chapter Four, that emotion is just reflecting variety that is already
there, this might make emotion at most harmlessly bad.

However, I will argue that a case can be made for making room for such variety. That this case is itself somewhat problematic points to certain deep-rooted antinomies between normative assumptions that in themselves seem unexceptionable. Specifically, in the problem cases, demands for universalisability, for moral responses to reflect their objects and not their subjects, and for agents to display not only moral goodness but also some measure of internal coherence *qua* agents, may all pull in different directions.

But again, it seems that emotion in itself would not be the source of such problems. Moreover, insofar as such antinomies seem to be part of what holds up a more fully-fledged positive account of emotion’s place in moral agency, it seems that the source of some apparent problems with emotion is antinomies in our metaethical thinking rather than problems with emotion as such. On the whole, then, a reasonably successful case for emotion can be made on general grounds: the problems are mainly about the particular forms emotions take. And insofar as there are problems with the particular forms emotion takes, blame can to a large extent be shifted off emotion and onto the complexities and contradictions inherent in our situation - and in morality’s attempt to overcome these.
## CONTENTS

*Abstract*

*Acknowledgements*

*Introduction*

### CHAPTER 1: AGAINST MORAL AGENCY WITHOUT EMOTION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter aims and some initial definitions</td>
<td>3</td>
</tr>
<tr>
<td>2) Why would we get rid of emotion? - possible benefits</td>
<td>5</td>
</tr>
<tr>
<td>Three kinds of possible benefits to removing emotion</td>
<td>5</td>
</tr>
<tr>
<td>Some initial reasons to be sceptical about stoic benefits</td>
<td>8</td>
</tr>
<tr>
<td>3) Emotion and moral status</td>
<td>9</td>
</tr>
<tr>
<td>The significance of participant reactive attitudes</td>
<td>9</td>
</tr>
<tr>
<td>Participant reactive attitudes, the objective attitude, and the stoic</td>
<td>11</td>
</tr>
<tr>
<td>Self and others in moral agency: More kinds of participant reactive attitudes</td>
<td>14</td>
</tr>
<tr>
<td>Tensions in the stoic position</td>
<td>16</td>
</tr>
<tr>
<td>Emotion and entitlement to moral consideration</td>
<td>18</td>
</tr>
<tr>
<td>4) Emotion and moral perception</td>
<td>20</td>
</tr>
<tr>
<td>Accounts of emotion and some meta-ethical approaches to moral values</td>
<td>20</td>
</tr>
<tr>
<td>Emotion, reason and moral perception</td>
<td>22</td>
</tr>
<tr>
<td>A case for emotion and reason in moral perception</td>
<td>23</td>
</tr>
<tr>
<td>5) Emotion and moral motivation</td>
<td>24</td>
</tr>
<tr>
<td>Emotion, reason and moral motivation</td>
<td>24</td>
</tr>
<tr>
<td>The case for the will</td>
<td>27</td>
</tr>
<tr>
<td>Interjection - desire and affect</td>
<td>28</td>
</tr>
<tr>
<td>Calm passions</td>
<td>30</td>
</tr>
<tr>
<td>6) Emotion, morality and flourishing</td>
<td>31</td>
</tr>
<tr>
<td>Hedonic tone and benefits of removing emotion</td>
<td>31</td>
</tr>
<tr>
<td>Morality and flourishing</td>
<td>34</td>
</tr>
<tr>
<td>7) Against moral agency without emotion</td>
<td>36</td>
</tr>
</tbody>
</table>

### CHAPTER 2: SOME WORRIES ABOUT EMOTION IN MORAL AGENCY

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Introduction</td>
<td>38</td>
</tr>
<tr>
<td>Chapter aims</td>
<td>38</td>
</tr>
<tr>
<td>Three kinds of charges against emotion</td>
<td>38</td>
</tr>
<tr>
<td>2) Objections to emotion from concerns with objectivity</td>
<td>40</td>
</tr>
<tr>
<td>Subjectivity as phenomenology</td>
<td>41</td>
</tr>
<tr>
<td>Subjectivity as brute feeling</td>
<td>42</td>
</tr>
<tr>
<td>Subjectivity as projection</td>
<td>43</td>
</tr>
<tr>
<td>Subjectivity as agent-relativity</td>
<td>48</td>
</tr>
</tbody>
</table>
Subjectivity and moral agency

3) Objections to emotion from concerns with autonomy
   Concepts of autonomy and concepts relevant for autonomy
   Emotion and autonomy
   Some worries about the value of autonomy

4) Objections to emotion from concerns with rationality
   Criteria of rationality
   Rationality in moral agency
   The cognitive rationality of emotion
   The strategic rationality of emotion
   Some possibly emotion-specific problems about rationality

5) Emotion and worries about emotion

CHAPTER 3: THE DIFFERENCE IT MAKES WHAT EMOTION IS

1) Introduction
   Chapter aims and outline

2) Judgementalism and the role of emotion in moral agency
   A strong cognitive account - the judgementalist identity theory
   Identity theory and the normative vindication of emotion
   Voluntarism about emotion and judgement
   The distinctness of emotion - just hot cognition?
   Some further discrepancies between emotion and judgement

3) Some ways we might want to revise the cognition-claim
   Strong versus weak senses of cognitive
   Cognitive scheme rather than state?
   Quasi-intentionality
   Cognition-claims and the place of the body in emotion
   The place of feeling in emotion
   The place of desire in emotion
   Weak cognitions and minimal rationality

4) The significance of the object-claim
   Object-appraisal and the cognitive rationality of emotion
   Object-appraisal and the strategic rationality of emotion

5) Beyond cognition and objects: Subjects, ideologies and emotion
   Ideologies of emotion
   Formalisation by subject
   Subjects, objects and grounds in emotion

6) Emotion and the revised cognitive account

CHAPTER 4: EXTERNAL SOURCES OF PROBLEMS WITH EMOTION

1) Introduction
   Some ways the problems might be inherent to emotion
### 2) External sources of the subjectivity of emotion

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epistemic fallibility</td>
<td>130</td>
</tr>
<tr>
<td>Perspective and relativity as sources of subjectivity - and a qualification</td>
<td>132</td>
</tr>
<tr>
<td>First-person perspective</td>
<td>135</td>
</tr>
<tr>
<td>Embodiment and (first-person) perspective</td>
<td>136</td>
</tr>
<tr>
<td>Perspectivity and intersubjectivity</td>
<td>138</td>
</tr>
<tr>
<td>Emotion and transcendence of perspective</td>
<td>139</td>
</tr>
</tbody>
</table>

### 3) External sources of heteronomy

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coming to recognise oneself and others as subjects</td>
<td>140</td>
</tr>
<tr>
<td>Vulnerability and dependency</td>
<td>142</td>
</tr>
<tr>
<td>The intersubjective as source of heteronomy</td>
<td>144</td>
</tr>
<tr>
<td>Sources of internal heteronomy</td>
<td>147</td>
</tr>
</tbody>
</table>

### 4) External sources of rationality-failure

### 5) Contingencies of outside sources and contingencies of emotion

Some possible ways to correct for bad contingencies - and possible roles for emotion in these...

### 6) Emotion and external trouble-sources

---

### CHAPTER 5: RE-EVALUATING EMOTION IN MORAL AGENCY

1) Introduction

2) Changing emotion and limits on change

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limits on the possibility of change</td>
<td>159</td>
</tr>
<tr>
<td>Limits on the desirability of change</td>
<td>160</td>
</tr>
</tbody>
</table>

3) Ideologies of emotion and ideologies of emotionality

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moral theories and ideologies of emotionality</td>
<td>162</td>
</tr>
</tbody>
</table>

4) How ideologies may transform emotions and some caveats

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elements of ideologies</td>
<td>171</td>
</tr>
<tr>
<td>Ideologies and &quot;natural&quot; emotion</td>
<td>174</td>
</tr>
</tbody>
</table>

5) Ideological shaping of emotion and some caveats

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideology as (mis)representation</td>
<td>178</td>
</tr>
<tr>
<td>Some points more specific to normative standards</td>
<td>180</td>
</tr>
</tbody>
</table>

6) The pros and cons of ideological adjustment

---

### CHAPTER 6: EMOTION, DIFFERENCE AND MORAL AGENCY

1) Introduction

2) Diversity and the objections to emotion

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjectivity and the subject/object-balance in emotions</td>
<td>186</td>
</tr>
<tr>
<td>Emotion, difference and internal coherence in agency</td>
<td>189</td>
</tr>
</tbody>
</table>

3) Does morality impose uniformity across agents?

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniformity - some usual (and one not so usual) suspect</td>
<td>193</td>
</tr>
<tr>
<td>Some reasons we might want to allow diversity</td>
<td>194</td>
</tr>
</tbody>
</table>
Some general pressures towards uniformity

4) Some possible ways to counter pressures towards conformity
   1) Deny complete determination 199
   2) Deny indiscriminate application 199
   3) Consistency between agents versus between cases 201

5) Some worries about the virtuous agent
   Costs of allowing patterns 203
   The unity of virtue and the unity of good 206
   A more flexible notion of character? 207

6) Emotion, difference and moral agency

CONCLUSION 211

BIBLIOGRAPHY 215
Chapter 1: Against moral agency without emotion

1) Introduction

Fairly ordinary experience of emotion suggests that emotion can be an influence for the worse. It may cloud and distort judgement. It may give rise to and feed unfortunate desires, such as vindictive ones. It may violate our notions of proportion and reasonable restraint. Or then again it may be too weak, the wrong kind, misdirected, or otherwise fail to fit the occasion. It tends to reinforce any accompanying beliefs and desires by amplifying them, giving them an added vividness of impact - which is obviously a further worry if those beliefs and desires are wrong ones. Once present, emotion also tends to be self-reinforcing. Angry, we have a sharpened awareness of occasions for anger, sometimes with the same target we first got angry at, sometimes broadening the range to the detriment of more or less innocent bystanders. Or again, emotion may be not so much wrong as random. It may come over us as if against or regardless of our will. And it makes us attach importance to, and be vulnerable to, things - and people - outside our control. Worse, these diverse worries can apply together. So emotion may not just be random, distorted, or distorting, and so on, but virtually any combination of them, at once.

Of course, all the assumptions above can be questioned. We might ask whether emotion really does these things, and if so, how emotion does them. We might wonder whether these perceived faults are due to emotion per se or are just unfortunate particular instances of emotion, and whether something other than emotion could be causing these effects. In some cases we might think that the ill effects attributed to emotion are not so bad at all - or that they are just flipsides of otherwise good sides of emotion. I will return to these issues further on in the thesis - for now, they will be worth bearing in mind.

The next chapter will set out charges typically raised against emotion in some more detail; their treatment in this chapter will tend to be fairly superficial. For
my immediate purposes, it is enough to show that there are *prima facie* reasons to think that emotion is bad for moral agency. The examples above suggest emotion may be detrimental to efficient agency generally, as well as to moral agency specifically. However, my argument will point in the opposite direction. My aim is to show that there are reasons to think that removing emotion - if this could be achieved - would be even more fundamentally crippling to agency, moral and otherwise.

*Agency* will here be understood throughout in a broad sense, to include internal mental states such as beliefs, desires, attitudes and of course emotions as well as more outward features such as actions and behaviours. Some more will be said on the subject of agency, and on connections and distinctions between moral agency and agency in general, here and in the next chapter. However, I will try to clarify these issues in the context of this thesis only as far as this is necessary and useful for the immediate purposes of the thesis topic. I will only try to get a clear enough sense of agency and moral agency - both difficult topics in their own right - to see why emotion has been thought to raise problems for moral agency.

**Chapter aims and some initial definitions**

For this chapter, the plan is, first, to outline some possible benefits of getting rid of emotion, and some of the possible costs. I will look at four kinds of possible benefits: Cognitive and motivational benefits, benefits to capacities for autonomy, and hedonic benefits. As against this, my basic argument is that getting rid of emotion is a "throwing the baby out with the bathwater"-solution as far as moral agency is concerned. Specifically, I argue that emotion plays an irreplaceable role in informing, motivating and even constituting moral agency. My main argument is from consideration of the crucial role of what Strawson calls participant reactive attitudes in moral agency, and the role of emotion in underpinning these attitudes. I also argue that there are more general reasons to think that emotion plays a crucial and distinctive role in moral perception and motivation. These arguments will also support the claim that emotion is crucial to participant reactive attitudes. And finally, I argue that losing the capacity for feeling emotions would diminish our
capacities for human flourishing.

**Introducing lower-case stoicism**

In the following, I will refer to the state of elimination or suspension of emotion as the stoic position. More specifically, I am interested in the view that such suspension or elimination of emotion would be beneficial to moral agency. This view will be referred to as stoicism.

The stoicism here is lower-case since it draws on colloquial use of the term "stoic" and is not meant to give anything like an exact picture of the position of the historical Stoics. The position still draws fairly heavily on the historical Stoics, and their influence on later thinkers. For instance, a number of the arguments against emotion that will be considered in this thesis have a recognisably Kantian flavour which may partly be due to Kant’s drawing inspiration from the Stoics.

However, some of the considerations that make lower-case stoicism tempting spring from other influences, such as utilitarian concerns with the way emotions underpin particular attachments that get in the way of utility-maximisation. The reason/emotion distinction that underpins much of the worry about emotion also precedes the Stoics, being notably present in both Plato and Aristotle, and can be traced (from whichever of these and other influences) in current folk psychology as well as in current philosophy.

---

1 For discussions of the historical Stoics on emotion, and their influence on later thinkers, see for instance Sorabji (2000), Lloyd (1978), Nussbaum (1994, especially Chapters Nine and Ten; Chapters Eleven and Twelve, also on Stoicism, deal specifically with Seneca). Sherman (1997: especially Chapters One and Three) and Schneewind (1996) are illuminating on the influence of the Stoics on Kant, and on the textual basis for emotion-negative views in Kant. On comparisons of Aristotle and Kant’s approaches, see again Sherman (1997) and also Annas (1996).

2 However, I think that although emotion may get in the way of utility-maximisation in individual cases, the kinds of goods that utilitarians aim to maximise, such as happiness, will not be possible if emotion is removed altogether. The utilitarian has good overall cause not to want emotion removed, even if emotion may cause trouble for their notion of good moral agency in a large number of specific cases. I will say more about this in the section on morality and human flourishing later in this chapter.

3 As the argument in this chapter will show, the implications of the reason/emotion divide are not straightforward. While the majority of those that distinguish sharply between reason and emotion give reason metaethical pride of place, there are also prominent examples, such as Hume, of writers who insist on a sharp distinction but give emotion pride of place. And as I shall argue in Chapter Three, how emotion is understood does not necessarily relate straightforwardly to whether it is thought to be meta-ethically “respectable”, either.
Emotion - postponing exact definition

A final introductory point: Clearly, what emotion is understood to be will be crucial in answering any further questions about the place of emotion in moral agency. While questions of definition are nearly always difficult, however, they seem to be particularly so in the case of emotion. Moreover, the kind of definition required - and the degree of precision required for it - will vary according to the reasons we are interested in what emotion is. So it should make sense to set out by establishing what the nature of our interest in emotion is in relation to the specific thesis question, and then return to the question of definition against that background.

Obviously, some assumptions will need to be made about emotion for discussion to get off the ground at all. And in fact, a number of assumptions about emotion have already been mentioned right at the start of this chapter; more will be added below. Insofar as these initial assumptions will tend towards defining emotion and its standing with respect to moral agency, I will try to make the assumptions, and their status as assumptions, as explicit as possible. It may not always make sense in terms of the flow of the argument to discuss the plausibility of an assumption as soon as the assumption itself is raised. In these cases, I will try to signpost as clearly as possible where this more detailed examination of the assumptions will take place.

2) Why would we get rid of emotion? - Possible benefits

What would be the benefits of getting rid of emotion? Considering the kinds of worries raised above, there are number of headings such benefits might come under. Here, I will concentrate on three main kinds: Cognitive and motivational benefits, and benefits to autonomy. A fourth kind, which I shall call hedonic benefits, raises broader issues about the place of pleasure and human flourishing in morality, and will be dealt with under the heading of issues of human flourishing at the end of this chapter.
Three kinds of possible benefits to removing emotion

Cognitive benefits

First, then, we have cognitive benefits. If emotion distorts the ways in which we try to capture truths about the world, then perhaps removing emotion will remove the distortion and let us see (literally and metaphorically) and reason clearly. Since emotion has been thought to affect not only perception in the broad sense, but also the validity of the inferences we make\(^5\), this would be a significant possible gain. This issue will be focused centrally and in more detail in Chapters Two and Three.

Motivational benefits - and introducing the concept of affects

Secondly, we have motivational benefits. Some of these come through removing distorting influences on perception and reasoning. The thought here is that if we saw more clearly than we do through emotion, or understood better how the fulfilment of emotion-related desire would actually turn out; we might also be motivated more wisely\(^6\).

Further, a number of morally problematic traits, such as selfishness, are usually thought to be affect-bound - constituted, causally brought about by, or reinforced by emotions\(^7\) - and by desires, feelings and other affects such as moods.

By affects I mean, very roughly, states that involve feeling, and the related dispositions to experience such feelings. Typical affects, I take it, will have hedonic

\(^5\) Aristotle, for instance, emphasises the role of emotion in leaping to conclusions. So at NE1149a24-b15 he says that anger may be like a servant who runs out before hearing the entire order, inspired "by reason of the warmth and hastiness of its nature" to rise up when argument or imagination presents cause for anger to us, without stopping for verification. And throughout book II of The Art of Rhetoric, he talks of the arousal of appropriate emotions as a short-cut to instilling in others the beliefs and attitudes the orator wishes them to have, and emphasises that the short-cut can work even where the orator's actual arguments are logically weak and flawed.

\(^6\) Cf. Hume 1978: 416: "the moment we perceive the falsehood of any supposition or the insufficiency of any means, our passions yield to our reasons without any opposition" - though this is possibly too much of a blanket optimist view, given the existence of "recalcitrant" emotions (of which more later - see also D'Arms and Jacobson, 2001).

\(^7\) Cf. for instance the central role in our psychological makeup that both Hume and Adam Smith (1759) assume for emotions that relate to what we would now call "self-image" and the regard of others, and the crucial emotional investment they both assume we naturally have in the advancement of our self-interest. Though of course both of these writers view such other-directedness with a degree of approval that a more Kantian autonomy-centred approach would not share.
(pleasant/unpleasant) tone, of which more later, though there may be more to feeling than that. While I shall not be setting out a general theory of affective states here, much of what is said about emotion will mutatis mutandis be applicable to other mental kinds insofar as they resemble emotions in involving feeling. Some of these other affect-types - notably mood and desire - will figure at various points in the discussion here and later in the thesis. I would argue that they have important similarities and links to emotion - but I will only pursue this here as far as a consistent main focus on emotion allows.

The removal of emotion, and perhaps of affect generally, might then reduce, if not altogether remove, the impact such traits have on motivation. Also, such character traits may be understood as being formed largely through the impact of emotional experiences, the traits then in turn causing, expressing, and reinforcing themselves through, emotion. And through emotion, these traits might then also influence thought, motivation and action.

There are also concerns about the way emotion affects our capacity for strategic rationality - that which concerns means-ends reasoning - that affect the question of emotion’s impact on motivation. For instance, counterproductive behaviour, including self-destructive behaviour, seems in large part to be continued through entrenched patterns of emotional reaction, and may typically originate in emotional reactions. I will have more to say about this issue of strategic rationality in Chapter Two, and more about emotions and traits in chapters Three through Five.

Benefits to capacities for internal and external autonomy

Thirdly, getting rid of emotion might be thought to increase autonomy, internal and external. I will have more to say on both counts in the next chapter - for now the following should do. By external autonomy, I have in mind issues like these: Being emotionally attached to what is outside any significant degree of control on our part - from the natural environment to the actions, attitudes and well-being of others - we have hostages to fortune. More, we ourselves are hostages - to

---

8 I would like to say, at this point, that affects involve some kind of pro- or con-attitudes towards intentional objects, but this point may be in need of some careful discussion given the supposed objectlessness of moods, and the fact that emotions are not always simply classifiable in pro- or con-terms. So I will postpone that part of the discussion till Chapter Three.
fortune, to events, and to the others whose actions and attitudes impact on us. And to the degree that such hostage situations obtain, our free agency is undermined. The more so, if emotion reinforces this hostage situation through the pain or pleasure of our feelings.

But emotion also undermines autonomy from within, insofar as emotions at least sometimes appear to come over us against or regardless of our will. This is particularly a worry since emotion, especially in its felt, occurrent form, tends to dominate our thoughts, perceptions, motivations and tendencies towards action. Insofar as it is something that we undergo or are passively carried away by, emotion might not be an expression of our agency. Rather, it features as a given about ourselves, as baggage or handicap for agency to work around - and so as a challenge to internal autonomy. Take emotion away, then, the stoic might argue, and no longer will our agency be hampered by hostage situations, or by emotional manifestations of psychological baggage or handicaps.

Some initial reasons to be sceptical about stoic benefits

Of course, the fact that something acts as a distorting presence need not mean that its absence would lead to optimal functioning. What remains after the clear-out may itself not be optimally functional - as far as emotion's role in moral agency goes, Hume's disparaging view of what reason can(not) achieve on its own springs to mind here (1978: 415). Also, that emotion can act as a distorting presence need not prove that it does not also make vital positive contributions - and I shall argue, below, that it does have vital positive contributions to make. Emotion may also have non-vital positive contributions to make, but that would hardly be enough to tip the balance in its favour - particularly not if its presence also has serious negative impact.

The general worry, then, would be that in removing emotion, we might lose what is good about emotion along with what is bad in it. Further, we might worry

---

9 Accounts of emotions as action-reasons go back to the very earliest philosophical accounts of emotion in e.g. Plato or Aristotle (on the latter, see e.g. Charles, 1984). For a Humean account of emotion as action-reason, see Baier (1985b).
that what is good in emotion cannot be replaced by other faculties\textsuperscript{10}, such as the use of reason alone. The crucial question is whether emotion has essential positive contributions to make, and specifically what relevance these have to moral agency.

Next, then, I will give four lines of argument in favour of a necessary role for emotion in moral agency. The first, as already stated, is from the crucial place of participant reactive attitudes in moral agency, and the crucial place of emotion in underpinning these attitudes. I argue that the stoic strategy of altogether eliminating or suspending emotion can be shown to be counter-productive and internally incoherent. In support of emotion specifically being crucial here, I also argue for a necessary and distinctive role for emotion in moral perception and motivation. And finally, I argue that our capacity for feeling emotion is crucial to human flourishing, and human flourishing to morality.

3) Emotion and moral status

The significance of participant reactive attitudes

First, then, I argue that emotion has a vital connection with moral agency through the conceptual connections between the applicability to a person of what P.F. Strawson (1974b) calls participant reactive attitudes, and ascriptions of moral agency. There are also arguable conceptual connections between the applicability of such attitudes and being held entitled to moral consideration - in my terms, being a moral subject. Insofar as recognising moral subjects is a vital moral skill, the connection between participant reactive attitudes and moral subjects will also be of some importance (again) in what follows. I will return to this issue briefly at the end of this section.

Moral agency

A very basic distinction in moral contexts is between entities that can be held

\textsuperscript{10} While I will use terms like reason and emotion (and further on, will and desire) as if they referred to faculties here, this is metaphorical shorthand for the sake of convenience and not an indication of allegiance to some kind of faculty psychology.
morally responsible and those that cannot. Moral agents, however otherwise defined, are entities that can be held morally responsible. The basis for distinguishing between the two types of entities is philosophically problematic on a number of counts, such as the free will issue, which will not be the focus here, but the distinction remains a fundamental and important one for most moral theories.

Moreover, for those entities that can be held morally responsible (moral agent will be used to cover this from here on), we make distinctions between areas that are relevant to their moral agency, and in which they can be held morally responsible, and areas where such responsibility does not apply. Fairly uncontroversial examples of the latter would be involuntary actions such as the knee-jerk reflex or blinking if someone waves their hand right in front of your eyes. Unfortunately, as soon as we move into more morally relevant areas, controversy sets in fast, and sometimes, there may be no clear boundaries between what is part of someone’s moral agency and what is not.

The specific ways in which criteria of appropriateness for liability to moral judgement are set out vary between different meta-ethical approaches, and will not be the issue just here. The point for the present argument is that such criteria appear to be at least coincident with the criteria for ascribing the status of moral agent to someone. They might also be taken to be identical with these criteria.

These criteria for ascribing moral status (in this case as agent) to someone are still distinct from the criteria for assessing their moral performance. For instance, the capacity for voluntary action might be thought to be one appropriate (if not necessarily necessary or sufficient) criterion for ascribing the status of moral agent to someone, and by extension their voluntary actions would, ceteris paribus, be liable to moral assessment. The criteria for assessing particular actions or reactions as good or bad would however still be a further matter for discussion.

Of course, in relation to emotion, issues of how we can be morally or otherwise responsible for what we cannot voluntarily control in the moment (or in some cases at all) arise with particular force. I will look further into these issues in the next chapter, when examining issues of autonomy. I will then add the qualification that what we cannot control may still reflect on our character as moral agents in significant ways. But I will also argue, there and in later chapters, that we may be more in control of our emotions, at least in the long term, than is usually assumed in either philosophy or folk psychology. For the immediate purposes of discussing moral agency here, though, these qualifications can be ignored for the moment.
**Moral status and the participant reactive attitudes**

Now, regardless of the particular moral theory applied, a consistent criterion for being a moral agent is being appropriately liable to being on the receiving end of particular kinds of reactions, such as resentment, gratitude, indignation, and so on. A moral agent is appropriately liable to be the target of such reactions only in respect of those areas that are considered part of their moral agency, whatever those may turn out to be. There is, then, at least a conceptual connection between the conditions of appropriateness for the attitudes involved in such reactions and ascription of moral agency. Similarly, there is a conceptual connection between being considered a moral subject and being a target for particular kinds of reactions such as sympathy, pity, or people feeling happy for or indignant on behalf of one.

The kind of reactions which moral agents are said liable to be the targets of, and those that apply to moral subjects, are both examples of what Strawson calls *participant reactive attitudes*, which are "the non-detached attitudes and reactions of people directly involved in transactions with each other" (1974b: 4) intimately or casually. To which we might perhaps add those indirectly involved in transactions with each other (O’Neill, 1996). Strawson’s own list of examples includes, as well as gratitude and resentment, forgiveness, love, hurt feelings - all affective or affect-involving attitudes.

**Participant reactive attitudes, the objective attitude, and the stoic**

Participant reactive attitudes also relate to what Strawson terms a "central commonplace": the great importance we attach to the attitudes and intentions towards us of other human beings, particularly those of significant others, and the great extent to which our personal feelings and reactions depend on or involve our beliefs about these attitudes and intentions, above and beyond manifest action (1974b: 5). The question then, in relation to the stoic, is whether they can get rid of

---

12 Cf. Nagel and Williams on 'moral luck' (Nagel 1979b, Williams 1981b).
13 But see Watson (1987), for some of the problems around applying liability to participant reactive attitudes as a criterion for responsibility.
14 Wallace (1994: introduction), argues that it is in fact this underlying attitude of engagement, rather than the participant reactive attitudes, that is constitutive of ascriptions of moral agency, though the distinction between the underlying attitude and its manifestation in the various participant
this attachment of importance to attitudes and intentions without at the same time suspending the participant reactive attitudes altogether. For if participant reactive attitudes have an essential connection to the recognition of others as moral agents and subjects, as Strawson’s account implies they do, this would be a problematic move for the stoic. The stoic may wish to get rid of emotion, and of emotions involving vulnerability to the actions and attitudes of others particularly, but they certainly would not want to get rid of any capacities vital to moral agency.

Now, the participant reactive attitudes are subject to variations in how, to what degree and whether they are natural, reasonable and appropriate in relation to situations and the agents involved. Some occasions for resentment, for instance, though they may invite suspension of occurrent reactions, do not invite the suspension of the reactive attitude towards the agent. If someone hurts us through ignorance, clumsiness or misguided good intentions, the presence of these mitigating factors does not invite us to suspend a general reactive attitude towards them. We do not stop seeing them as moral agents: we merely extend goodwill to them by suspending sanctions, given that their intention was not such as to invite justified resentment. In this case, we inhibit resentment without inhibiting the sort of demand that generally gives rise to resentment when unfulfilled. (Strawson 1974b: 7)

In other cases, however, the participant reactive attitude towards the agent may be suspended, either temporarily (“He was not himself”) or more permanently (“He’s not normal; what can you expect?”). By contrast with the participant reactive attitude towards agents, Strawson calls this attitude to agents the objective attitude. The objective attitude involves suspending ascriptions of moral agency. This is not to deny that we may still feel considerable distress at the agent’s actions, but only to say that the distress we feel will not be a participant reaction. Viewed objectively, the other qua agent occupies much the same kind of moral status, or rather non-status, as children and animals are usually assumed to have. He may on the other

attitudes, is a fine one. This distinction roughly parallels the distinction I will make in this thesis between the general capacity for being affected and the particular forms through which this capacity manifests itself. However, I shall argue that while the general capacity is necessary for the connection to moral agency, it is not sufficient: Someone whose attitude of engagement was lacking in recognition of others as persons with rights of their own would for instance be deficient in the form of their participan reactive attitudes, even if they were genuinely engaged in how those others and their attitudes affected the agent’s own interests.
hand still be a moral subject, since his being liable to be held responsible or not need have no bearing on ascribing that status to him. (1974b: 7-8)

A consistently objective attitude, though it does not preclude interaction with others, precludes participation in full interpersonal relationships, including the more abstract and impersonal ones: "If your attitude towards someone is wholly objective, then though you may fight him, you cannot quarrel with him, and though you may talk with him, even negotiate with him, you cannot reason with him. You can at most pretend." (1974b: 9).

The benefit of the objective attitude is that it offers a refuge from the strains of involvement, and so it may be invoked towards others "for reasons of self-protection, as an aid to policy" or simply out of intellectual curiosity (1974:9-10). Or even towards oneself, though Strawson has relatively little to say on this score: There may clearly be much relief in the suspension of liability to what he calls self-reactive attitudes (e.g. pride, shame, guilt, and sense of obligation or responsibility). As a consistent strategy, however, the objective attitude towards others comes at the cost of personal isolation: the objective attitude hinders (potential) resentment by hindering the establishment of any genuinely interpersonal relationships. Relationships of some kinds may still be possible in such cases - owner-pet relations and to some extent child-adult relations work along such lines, as do interactions and relationships between normal human agents and those who are considered sub-normal. But these relationships and interactions will be qualitatively different from those that normal human adult agents have with each other.

Now there does not seem to be anything immediately logically contradictory in universalising the objective attitude across all persons and all time; this is in fact exactly what a stoic argues for. Strawson however states himself to be "strongly inclined to think that it is, for us as we are, practically inconceivable" (1974b: 11): a consistent and universalised objective attitude is not compatible with the background

---

15 The kinds of policies that Strawson seems to have in mind are social policies for managing agents, e.g. through social, legal and political sanctions.
16 On the significance of self-reactive emotions, see especially Taylor (1985).
17 Special cases, however, are those where we engage with others towards whom we take an objective attitude, but do so with an explicit view towards aiding them in becoming the kind of agents towards whom we can have participant attitudes. Adults, especially parents or teachers, might have this kind of relationship with children, and therapists with patients. (1974:19-20)
assumptions we normally make in moral reasoning.

If Strawson’s arguments hold, then, the stoic strategy may be in serious trouble, since it essentially involves the consistent suspension at least of participant reactive emotions, but seemingly also of participant reactive attitudes towards all others than the agent’s self. (By participant reactive emotions, I mean the emotional reactions associated with participant reactive attitudes). Insofar as the participant reactive attitudes dispose us to the occurrent reactive emotions, the stoic may be the more inclined to suspend the attitudes so as to cut off the attendant emotional reactions at the source. One consequence of this is that other agents cease to be agents, at least in a moral sense; more on this shortly. This problem ties in with the more far-reaching one implied by Strawson’s arguments: if participant reactive attitudes are in some sense essential in informing, motivating and constituting moral agency, how will the stoic manage as a moral agent without them?

The stoic position and the objective position - some differences

It is however worth making clear at this point that the stoic position is not simply the objective position. For the objective position involves the suspension of participant reactive attitudes only: one might take a consistently objective attitude towards all other putative moral agents and still be liable to emotions. As Strawson says: “The objective attitude may be emotionally toned in many ways, but not in all ways: it may include repulsion or fear, it may include pity or even love, though not all kinds of love” (1974b: 9). And we may still be affected emotionally by happenings that are not usually attributed to agency of any kind - natural disasters (barring theodicy issues) being a case in point.

So the stoic, if he is to succeed in eliminating emotion, must also get rid of those emotional reactions that remain after participant reactive attitudes have been suspended. Some of the problems raised by this will be discussed below: emotions of a non-participant-reactive kind may for instance play a considerable part in most conceptions of human flourishing. Understanding our own emotions of this kind, and those of others, seems a vital moral skill, both for doing good and for avoiding doing ill. Insofar as the stoic position goes beyond just suspending participant reactive attitudes, it is also liable to a range of worries about what it would mean for
our understanding of morality to erase from the picture the full richness of our actual emotional natures.

Self and others in moral agency: More kinds of participant reactive attitudes

Related to the kinds of participant reactive attitudes already mentioned are sympathetic and vicarious attitudes. We make moral demands not only on our own behalf, or of others, but also for others of others, and of ourselves for others. To which can perhaps be added the demands we make for ourselves of ourselves (Vetlesen 1994: Chapter Four): We could call these reflexive participant reactive attitudes. Failing to act right towards others, one might feel one had let oneself down as well as them. One might also have more specifically self-related undertakings, trivial or otherwise, towards self-improvement.

Vetlesen also uses this idea of demands for oneself of oneself towards formulating a normative concept of self-respect where “being-towards-oneself” is argued to be foundational to the project of constituting oneself as a moral agent and subject in relation to others. He suggests this modification brings Strawson’s account close to a Kantian notion of moral agency - seeing oneself non-instrumentally, and constituting oneself as a citizen of the kingdom of ends. In other words, the kind of attitudes that constitute the agent internally as a moral agent might be a special form of participant reactive attitudes.

These various types of reactive attitudes are connected, according to Strawson, “not merely logically, but also humanly”: “In general, though within varying limits, we demand of others for others, as well as of ourselves for others, something of the regard which we demand of others for ourselves” (1974b: 14-15). A normal moral agent, says Strawson, will display all three forms of reactive attitudes - or all four forms, if we concede Vetlesen’s point. To see oneself as the only moral agent would be a form of moral solipsism, but this is “barely more than a conceptual possibility, if it is that” (1974b: 15). Other variations would yield equally odd moral agents: saintly characters making demands only of themselves for others; agents who make demands only of others for others; radical moral solipsists who make demands only of themselves for themselves. Equalisation of moral demand,
Strawson emphasises, is meant to correct for the distorting effects of self-interest. The point of developing other kinds of reactive attitudes than those involving demands of others for oneself is to overcome self-centredness, not merely to replace it with different and more exotic distortions.

The oddness of the abnormal moral agents has a logical-conceptual basis. If I am a moral agent and others are moral agents, then qua moral agents the same basic demands would hold on us all, and to the same extent. In view of the possibility of suspension of participant reactive attitudes, the equalisation would seem to be a ceteris paribus principle: if in any particular case there is reason for suspension of participant reactive attitudes, then for that particular case the assumption of equality will be modified. Such modifications would however seem to require underpinning in the form of some relevant difference between cases where the participant reactive attitudes are suspended and those in which they are not.

**Tensions in the stoic position**

**Self/other asymmetry**

The difference on which the stoic approach effectively draws, however, is that between self and other. In practice the stoic approach amounts to a form of solipsism about moral agency: I can make demands only of myself. Others may still be counted as moral subjects, whom demands can be made for: the stoic is not an egoist, in any sense of that term. In fact, the asymmetry goes the other way: stoics make demands of themselves for others, but not of others for themselves. The self-other asymmetry is further reinforced by the strong perfectionist tendency in stoicism - a tendency implicit in the notion that moral agency would improve if we could just eliminate those pesky emotions. This perfectionist tendency also inclines the stoic to make strong demands of himself for himself. For perfectionism involves strong demands on moral agents, which for the stoic is essentially a constituency of one.

Whether a stoic can, without internal inconsistency, make claims of others for others is unclear, since the stoic can, apparently, demand for others, at least of himself, but not of others. However, the stoic's most immediate motivation for
suspending participant reactive attitudes is the loss of autonomy that goes with emotional involvement. As Strawson suggests, taking the objective attitude from a vicarious position may be psychologically easier - we are perhaps more inclined to take an objective view of others where we are not the ones directly engaged with them or affected by them. However, in the vicarious case, the motivation to suspend reactive attitudes is weaker, as there is less antecedent involvement to seek refuge from. Nevertheless, this possibility would leave the stoic at least potentially vulnerable to having vicarious participant reactions.

The possibility that stoic agents can make demands of others for others, but not of others on their own behalf, also brings out a further consequence of the stoic's wish to avoid involvement. The stoic needs to see himself or herself, and no-one else, as a moral agent, but only others as moral subjects. This will have the benefit, for the stoic, of precluding other-reactive participant attitudes on the stoic's own behalf. No matter what distress others may cause the stoic, it will not be further exacerbated by reactions such as resentment, disappointment or sense of betrayal. And of course, the stoic also intends to remove the non-participant-reactive emotions, distress included. By losing feelings on his own behalf, the stoic has then made a considerable gain in peace of mind, even if he is still vulnerable to emotional reactions on others' behalf.

A main problem, then, is whether the self/other asymmetry can be justified - especially if we do not share the stoic's view that avoiding the pangs of participant reactive attitudes is a practicable and worthwhile aim. For this motivation aside, what could justify the asymmetry? While not typically self-seeking, the stoic in this respect finds himself in the company of the ethical egoist, treating the self/other-distinction as a foundational moral distinction18.

Where there is Kantian influence behind the stoic drive, the stoic might also be troubled by the fact that his effectively withholding the status of moral agent from others signally fails to show the respect for others qua moral agents inherent in the notion of the moral project as a kingdom of ends. It is one thing to withhold such

18 For more general discussions of the meta-ethical issues raised for moral agency by self/other asymmetry, see for instance Slote (1984). Stocker (1976a, 1976b, 1997) defends self/other asymmetries in moral agency, but not on the kind of lines taken by the stoic.
respect where it is clearly not warranted, or where it simply fails to apply, as one might think is the case for children and animals. But to make oneself the only moral agent where others have a *prima facie* equal claim to the same status is rather more worrying.

*Internal inconsistencies in stoicism*

The stoic might then, for the sake of consistency, find himself driven in the direction of universalising his claim so that each agent will assume the same foundational moral distinction between himself and others. Each individual stoic agent, then, will believe in the existence of other moral subjects. In fact, if the argument above holds, he will only believe in other moral subjects, since he is not himself one. But he will be solipsist about moral agency - while at the same time holding that everyone else should take the same solipsist view, so that he must acknowledge that he himself is, from anyone else’s point of view, a moral subject but not a moral agent. And such universalised solipsism hardly seems a happy outcome in itself. We might also want to apply, at this point, Strawson’s argument that “it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition” (1974b: 11).

Interestingly, self/other asymmetry and solipsistic tendencies (tendencies, that is, to treat other people as less *real* people, and as lesser moral subjects) are often cited as typical features of emotion and as one of the reasons why emotion is not morally and philosophically respectable. This is particularly ironic given that the stoic’s strategy was aimed at removing emotion from moral agency. Which may go to suggest that the problem of self/other asymmetries is not a problem about emotion as such: rather, self/other asymmetries built into our existence as separate individual units of agency are reflected in emotion. (I shall have more to say about this in Chapter Four.)

At this point, moreover, it seems that the stoic is still relying on participant reactive attitudes to inform his moral agency. Admittedly, he has suspended reactions on his own behalf, and towards others, but it is not clear that he has altogether moved away from having participant reactive attitudes. Instead, he merely has a rather skewed version of them, due to the asymmetry his strategy imposes.
between himself and others. So far, then, he emerges as less than fully successful in suspending participant reactive attitudes. He is also, in several respects, caught up in an internally inconsistent position. For the participant reactive attitudes that he aimed to remove are still needed to inform his moral agency, so it seems they cannot have been genuinely removed. He also treats otherwise like cases differently on grounds (such as self-other asymmetry) that he himself might be expected to reject elsewhere, and is even using as part of his argument for removing emotion. These binds look particularly bad since the stoic aims to be a good moral agent - that, after all, is what he sacrificed emotion to achieve.

**Emotion and entitlement to moral consideration**

To these considerations of how emotion might be necessary to moral agency, we can add that on some meta-ethical accounts, the capacity for affect is thought constitutive of entitlement to moral consideration. For instance, classical utilitarianism takes hedonically toned sentience\(^\text{19}\) both as an indicator of value and as constitutive of value subject to the quantitative restrictions of utility-maximisation, in either case, but still. Affects such as emotions are prime value-makers and instances of our hedonically toned sentience. And it is as bearers of such moral value-makers that we are, on this theory, entitled to moral consideration. And the existence of moral subjects, so understood, is, to the utilitarian, what gives moral agency its point, and what defines it. That we are producing happiness is both the way in which we do good, and the criterion of whether we are doing good\(^\text{20}\).

Given this, the capacity for affect also seems crucial for moral agency insofar as it is necessary (if not sufficient) for empathy with other moral subjects. If we did not ourselves understand feelings, we would probably not be able to recognise them

---

19 I take it that sentience in itself need not be taken to entail capacity for feeling, or hedonic tone of any kind. Cf. also the conceptual overlap, and the overlaps in usage, between "sentience", "(self-)awareness" and "consciousness", the two latter of which also need not be taken to involve capacity for feeling. Svendsen (1999, Chapter Four) argues on Heideggerian grounds that all intentional awareness may have at least "background" Stimmung, which he understands as something like affective tone, but we need not make quite this radical a move to make room for affective tone to be crucially important.

20 This approach to morality of course precedes utilitarianism - precursors of it can be traced right back to Hellenistic philosophers such as the Epicureans, and of course to Aristotle, as well as to relatively more recent sources such as Hume, as well as to Bentham and the Mills.
in others, and so might miss crucial moral information. And not only would we not be capable of such empathy if we did not ourselves have feelings, but insofar as empathy itself involves feeling, we would be doubly handicapped. Similarly, capacity for feeling emotion seems required for sympathy. And for in-his-shoes-imagining of the situation of the moral subject we are dealing with, we would presumably also require emotion-related capacities. Merely imagining the other’s situation would not get us very far as regards informing or motivating relevant actions if we could not also at least imaginatively engage with their emotional experience. 

I do not want to argue that capacity for affect is a necessary requirement for being a moral subject, nor that it is the only requirement. But I do think that capacity for affect is a crucial, and sufficient, basis for being a moral subject. Nor need we adopt an exclusively consequentialist hedonist approach to concede that pleasure and consequences may be issues of moral relevance for human agents, given the way we happen to be constituted.

So the capacity for feeling emotion may be crucial both for our status as moral subjects and for our capacity to perform well as moral agents. In the next two sections, I examine further why emotion specifically should be crucial to informing and motivating moral agency in this way. In the final section, I argue for the crucial role of emotion in human flourishing, a theme that picks up again on some of the arguments just made about moral subjects.

4) Emotion and moral perception

Accounts of emotion and some meta-ethical approaches to moral values

In looking into the connection between emotion and entitlement to moral consideration in the previous section, my argument suggested both that emotion is constitutive of moral concerns applying at all, and that emotion has an essential and

---

Mackie (1977: Chapter Four), Wollheim (1984), Vetlesen (1994) and Goldie (1999) present somewhat more fine-grained accounts of different ways in which we understand others' emotions, an issue I shall return to in Chapter Four. For immediate purposes, it seems enough to give some further
distinctive role in moral perception. In this section, I will consider some further arguments in favour of these views. However, I will not be arguing that emotion is necessarily sufficient on either score, or that it is infallible. Instead, I will sketch a case for reason and affect in morality. While I draw on considerations about the distinct nature and functions of affect and reason, I will argue that the two faculties also overlap, and that this overlap may itself serve useful purposes.

The possible benefits of removing emotion mentioned above included cognitive and motivational improvement. Whether this convinces also depends on what assumptions we make about the nature of moral values and value-perception, and the implications such assumptions have relative to emotion.

For instance, if we assume that rational thought is sufficient for correct moral perception and judgement, then emotion is at best superfluous and at worst an impediment. On the other hand, if we take a more Humean approach, and assume that evaluative judgements of any kind are ultimately based in emotional reaction and impossible without it, then it seems moral agency without emotion would be at best severely handicapped and at worst entirely impossible.

This is of course in Hume's case supported by additional assumptions about the radically distinct nature of reason and emotion at the psychological and epistemological level, and the equally radically distinct nature of fact and value at the ontological level. Reason, on this account, is concerned with matters of fact and their relations only, and is motivationally inert (Hume 1978: 457, 458). Its role in moral agency is and should only be instrumental (1978: 415). Value-perception on the other hand is the realm of passions (emotions and desires), which do have motivational force.

To this we could add that Humean subjectivism about value means that for value, ontological and epistemological levels are not very strongly distinct. Emotional reaction is not just ultimately what makes perception of values possible, but is also constitutive of value - good and bad are not in the things that we attribute them to, or in non-evaluative facts about them.

---

initial plausibility to the claim that emotion might be necessary for moral agency.

For some recent attempts to work out a Humean account of the ontology of value, as well as a Humean theory of value-perception, while at the same time avoiding radical moral irrealism or non-
Which is of course not to say, at least not in Hume's case, that anything goes where value-judgements are concerned. Restrictions based on whether there is interpersonal agreement of specific kinds about the judgements are for instance applicable, hence the importance of the general point of view\(^{23}\). The restrictions on what count as valid value-judgements cannot however, on this account, be ultimately based in non-evaluative facts. This is obviously not to say that non-evaluative facts might not be *morally relevant* facts\(^ {24}\). Obviously certain kinds of natural features of an entity or action might be thought relevant to judging it good or bad, but the goodness or badness does not lie in those facts themselves.

Now if emotion, or affect more generally, is necessary for moral values to obtain, then the notion of moral agency without emotion seems not only impossible but also absurd. As a meta-ethical position, however, this Humean view is itself rather controversial, and it falls outside the scope of this thesis to defend it, or even to examine the arguments for and against in much detail.

**Emotion, reason and moral perception**

Of course, one need not assume that emotion is ontologically constitutive of value, moral or other, to think that it necessary for value-*perception*. One might reject ontological sentimentalism about moral value and still insist on the epistemological part of the account. Reason alone might not reveal moral value even if value were objectively in the world\(^ {25}\) and not just in the mind. Even if emotion were only necessary for us to *know* moral truths, but not for them to obtain, the effect on moral agency of removing emotion would still be devastating. If other faculties still allowed us some degree of epistemic access to moral facts, we might perhaps be able to compensate to some degree. But reason as conceived by Hume...
would seem to be clearly under-equipped for the task.

For instance, if we allow that reason, so understood, could pick up on what kind of facts were morally relevant, and further assume that there is some pattern to moral truths that reason could uncover by applying formal rules or principles, then perhaps reason could manage in the absence of emotion. But if emotion is also what reveals certain kinds of fact as morally relevant in the first place, reason can only manage by itself once this revelation has been provided. So it seems we could do without emotion, but not right from the outset.

And two further considerations raise doubts as to whether reason would be up to the task even then. First, a particularist might argue that there is no such a priori pattern to correct moral judgement. Secondly, even if there were such a pattern, emotion might be needed in an ongoing way to reveal it. If emotion, as a mode of awareness, is sufficiently different from reason that its deliverances cannot be translated easily and without some loss of content into for instance beliefs, then perhaps the sensibility that we have as emotional beings cannot be adequately mimicked by reason even in an after-the-fact way.

Some philosophers, particularly those of a Kantian bent, have of course held that reason by itself can provide such revelation, and that emotion, far from being essential to revealing value, is more likely to be misled by appearances, by its own nature, or both, into wrong evaluations. Such accounts typically tend to view values as in some sense objective (whether in a realist or a cognitivist sense) and their perception by reason as an instance of reason’s revealing facts about the world to us. Which is not to say that the way in which reason reveals value is necessarily assumed to be the same as the way in which it reveals other facts, or that moral facts (if such exist) are facts in the same way as natural facts: on such points accounts

27 Continuing the loose Star Trek theme of current literature on emotion, the android character Data (in the second Star Trek series) has an emotion -chip which can be turned on and off at will, allowing him to stop or start experiencing emotional reactions to situations as desired. This might seem an enviable capacity. However, if I am right that it may not be a priori predictable even to a reasonably experienced agent what the correct emotional response to a situation is, or even whether it warrants emotional reaction, Data’s capacities would come with their own risks.
28 As mentioned in the section on possible cognitive benefits of removing emotion, above, such a view of emotion is also held by Aristotle, who still gives emotion a crucial role in moral perception as long as it is in accord with reason and can be brought under its control.
may vary. This assumption, I think, runs into more trouble when applied to motivational issues than when simply talking about moral perception.

However, it also seems plausible that moral perception may only be possible against a background of certain kinds of affective engagement with the world, through emotion or desire or both. And in fact, even Kantian accounts take some emotions, such as respect for persons, as basic to the moral point of view. I have argued, above, that a whole range of emotion-types, namely those involved in participant reactive attitudes, may be essential to constituting our moral agency. If an affective attitude of engagement is generally necessary to underpin participant reactive attitudes, we can then begin to see how emotion could be necessary to moral agency.

A case for emotion and reason in moral perception

As already stated, I do not want to argue for emotion as being sufficient for correct moral perception, only that it is necessary for it. My preferred solution is a mixed account of moral perception, with both reason and emotion (and possibly also desire), having parts to play. The question of the division of labour, and the internal ranking, of the faculties involved then arises.

What I have in mind is roughly this: It would seem rather pointless to have two qualitatively identical or near-identical faculties doubling up the moral work. If, however, there are two qualitatively different moral faculties, say one of affective awareness and one of critical reasoning, and both make distinct and useful contributions, it would make sense to have two faculties. But if one of them only deals with evaluations, and does not itself pick up on facts, and the other deals only with facts and cannot do evaluation, to give a rather crude version of the Humean picture, then either one without the other would be helpless.

Better, I think, to allow some degree of doubling up as well as complementary contributions. If affect involves some cognitive or representational element, and reason can at least get a rough handle on what the deliverances of the affective element would be, then we may hope that deficiencies in one faculty could be compensated for by the other. Then, for instance, we might be able to recognise
cause for compassion even in the absence of feeling, and conversely, to feel compassion even where reason fails to register cause.

Of course, cases where the two faculties come up with different answers would not be desirable from the point of view of coherent agency: that way lies 

\textit{akrasia} and struggles for self-control. But it needs to be conceded that emotions can go wrong. More controversially perhaps, it needs to be conceded that reason can go wrong. The actual reasoning capacities of real agents are often poor things compared with the transcendental grandness of capital-R Reason as conceived by Plato, Aristotle and Kant\textsuperscript{29}. Wrong beliefs and flawed inferences abound, and we may not be able to chalk up all these flaws to the pernicious influence of affects. If we give up the idea that truly good moral agency must always be reason-led\textsuperscript{30}, then we can make a case both for the necessity of emotion \textit{and} for the view that the possibility of \textit{akrasia} is, informally put, a feature rather than a bug\textsuperscript{31}. Of course, we cannot rule out the possibility that both faculties could go wrong at the same time, but the chance seems somewhat slimmer if the two can at least theoretically be got apart. I will leave this argument in this somewhat sketchy form here, but come back to it in chapter 3, in discussing the difference it makes what emotion is.

\textbf{5) Emotion and moral motivation}

\textbf{Emotion, reason and moral motivation}

Perception apart, there is also the question of how emotion features in moral

\textsuperscript{29} Though Plato and Aristotle, unlike Kant, are fairly happy to take it for granted that most people will not achieve this ideal.

\textsuperscript{30} In this, Aristotle and Kant are arguably more similar than is usually thought (cf. Sherman 1997, 1998, Broadie 1991) Both are rather disparaging of the value and tenability of \textit{natural} virtue, as manifested in spontaneous right feeling and action, and view true virtue as grounded in a state where the passions follow the lead of reason. Kant's account differs from Aristotle on this point mainly in seeing the assent of passion to reason as ideal but not necessarily required for virtue.

\textsuperscript{31} Explicit arguments to this effect can be found in Bennett (1974), Jones (2001); fairly clear suggestions of similar views can, I think, be found in Blackburn (1998, Chapter Four). See also Ben-Ze'ev (1987, 1997, 2001). Insofar as the argument above suggests a useful aspect of the distinctiveness of emotion, Damasio's arguments for the neurophysiological underpinning of such distinctness, and its implications for agency (Damasio 1994, 2000) may also be relevant here, though Damasio seems to operate with a rather more Humean account of rationality than I have suggested in this section.
motivation. The Humean account, as usually understood, assumes that reason is motivationally inert. Moreover, it assumes that the kind of facts which reason is capable of uncovering (i.e. non-evaluative ones) are also, in themselves, not motivating, or at least not necessarily motivating. And this has the effect of making passions doubly necessary to moral agency. Not only can reason not reveal value to us, on this account, but it cannot move us, either - not of its own accord and not through the kind of facts about the world that it reveals.

I do not want to argue that reason has no role in moral perception - in fact, I argued above that reason, as well as emotion, may have a necessary role. But if, as Aristotle suggests, reason only points us in the right direction but does not move us there, then we will need something other than reason to make us act, and most of the obvious candidates, emotion and desire notably, are affective states. And while the worries that apply to emotion are not necessary all going to apply to other kinds of affective states, other kinds of affective states - desires and moods, for instance - have their own long-standing normative concerns associated with them. And these worries are along much the same kind of lines as for emotion: that they are projective, that they distort perception, that they are not under our control; that they threaten autonomy, and so on.

So it seems that if we do not want moral perception as performed by reason to be motivationally inert, we must either reject the view that reason is in itself motivationally inert, or we must associate with reason some faculty that does

32 This is one way of generating the distinction between ethical internalism and ethical externalism. However, since motivating force has, plausibly or not, been attributed to reason by some authors, the internalism/externalism debate does not rest entirely on the issue of whether emotion or reason is the basis of moral capabilities. A "sentimentalist", who takes emotion to be the psychological basis of moral agency, may have less of an uphill struggle ahead of them to get to a plausible version of internalism, given that emotion, unlike reason, is usually thought of as inherently motivating. But they might be hindered by examples of cases where agents fail to have the proper emotional reactions, and so by extension to be emotionally motivated to do the right thing. (Blackburn (1998: 61) suggests that while externalists may win individual battles on this, the examples they rely on are basically parasitic on a background connection between value and motivation. And so the externalist cannot win the war: "the fact of a joint being out presupposes a normal or typical state of in which it is not out").

33 Blackburn (1998:90-91) casts the distinction between reason and these less respectable elements in Nietzschean terms, comparing the discovery that reason, due to a motivational deficit in itself, might need to rely on affects for moral motivation to the discovery of a Dionysian frog at the bottom of the Apollinian well.

34 "NE 1139a22-b11".
possess motivational force, but which also hopefully looks more respectable than emotion or desire. Or we must review our view of emotion and desire.

Failing such revision of the moral psychology, we risk getting a picture of moral agents as possessing the capacity for moral perception through reason, but with their capacity for moral motivation left to faculties, such as emotion, whose reliability has been called into question, and whose links with reason have been assumed to be tenuous. On this picture, whether we are capable of being motivated by our moral perception or not seems to be down to luck.

Now admittedly, this might be the case. Nor would taking a pessimist view of our capacity for being morally motivated be obviously inconsistent with an insistence on the rational nature of moral perception. And we might add into the pessimist picture that not all of us are capable of that, either, and certainly not consistently.

There does however seem to be a tendency, also on the part of those who think moral perception is done by reason, to resist fully embracing a pessimistic view of human moral psychology in this way. Some rationalists, like Plato, embrace an elitist approach: moral virtue is possible, but only for the few. But others, like Kant, want to insist that the capacities required for moral virtue, including right motivation, are at least potentially universal.

Attributing motivational force to reason itself has been thought problematic. This might just be the legacy of Hume’s influential conjoined distinctions - reason/fact/perception versus emotion/value/motivation. But it is notable that attributing motivational force to reason is already viewed as problematic by both Aristotle and Plato, even though the latter at least seems to affirm clearly that reason does have such motivating force.

The exact reasons why it is problematic to attribute motivational force to reason are not that easy to get at, and fall rather outside the scope of this thesis. It is not simply that reason deals with matters of fact and that evaluation (of any kind) is thought by some to go beyond mere fact, as suggested by Hume, and by Moore’s

---

35 Plato, in The Republic, notoriously took this as a basis for suggesting that these few should rule the many.

36 For instance, all three parts of the soul distinguished in The Republic - reason, the “spirited”
open question argument. There is also a distinction, and a potentially real conceptual and psychological gap, between evaluation and the motivation to act on the evaluation. Where emotions, desires or both are doing the evaluative work, we might expect the gap to close: *ceteris paribus*, we tend to expect that emotions and desires are moving forces.

**The case for the will**

The remaining candidate for motivational force, short of attributing motivational force to reason in itself, is usually the will, a faculty which also suffers from lack of definition, both as to its nature and as to its relations to other faculties. At the most basic, will may be defined as our capacity to make and hold to decisions. What informs those decisions, however, may be entirely open to debate, and the will usually does not appear to have a content of its own, apart from what reason or affect gives it. It seems will may diverge from reason, operate independently of it, or even (if we want to get really Nietzschean) act on or determine it, informed by motives that are themselves typically affective. So for instance Hume (1978: Book Two, Part Three) gives an account of will according to which its influencing motives are the passions, and repeatedly denies that reason can ever be any motive to the will. Will on Hume's account resembles the passions also insofar as it is not capable of being true or false, reasonable or unreasonable, in itself, but he says it is not strictly a passion and cannot create new sentiments part and the appetites - are thought to have motivational force each of their own (Plato, 1941).

---

37 Cf. for instance the introduction to Darwall, Gibbard and Railton (1997).

38 One *ceteris paribus* qualification is that the motivating force may be impeded by demotivating factors such as opportunity cost or the presence of other, contradictory, urges and drives, or mere inertia. It is also worth noting that while desires are usually assumed to be motivational by definition, the connection between emotion and motivation is not always straightforward. Some emotional reactions, such as surprise, awe, nostalgia (possibly for different reasons) have little inherent motivational content, though in this they seem atypical. More commonly, there may be some degree of under-determination as to what desires or desire-like states are entailed by being in a particular kind of emotional state.

39 On the meta-ethical status of the will more generally, see e.g. Dilham (1984), Clark (2000).

40 Cf. the common identification of *akrasia* with weakness of will - though Aristotle notably generates the problem of *akrasia* without introducing volition into the picture. See also Holton (1999) for an argument that weakness of will and *akrasia* are in fact distinct phenomena.

41 This point is crucial to both Nietzsche's theory of exercises of reason as being manifestations of the will to power, and relatedly to his theory of *ressentiment*. (Cf. Nietzsche, 1887, Essay One.)

27
It is not obvious, then, that will would necessarily do better in relation to reason, or to morality, than emotion or desire. And even where, as seems to be the case in Kant, will is subsumed under reason, as a capacity or faculty of reason, it tends to bear a suspicious family resemblance to affect, if only in terms of its functional role. Like desire, and perhaps emotion, will seems to involve a drive towards, not mere contemplation. Like desire, its direction of fit seems to be from agent to world rather than vice versa. Moreover, it seems doubtful, given the notorious and recalcitrant problem(s) of free will, whether will would do any better than emotion or desire as far as autonomy goes: will need not be free or autonomous will. (More on this in the next chapter).

When it comes to moral motivation, then, even accounts of moral agency that see reason as by and in itself sufficient for moral perception seem to import into moral agency something that, if it is not actually affective, resembles affect in at least some respects. Add to this that at least some kinds of affect-related attitudes, such as respect for persons, are typically insisted on as basic to informing adequate moral agency, even by writers as critical of affect as Kant. It seems then that at least some aspects of affect, or some affect-like features, will be crucially necessary for moral agency, however otherwise problematic affect might be.

Interjection - desire and affect

I have talked, above, of desire as an affect, which might seem a strange and unnecessary idea. While I will not be relying on it for my main argument here, it seems worth noting a few points on this issue. This also allows me to lead into a more general point about what Hume calls calm passions that may help motivate a more positive view of affect in moral agency.

Historically - and until fairly recently - desire tended to be classed with the passions, at least as a related kind of phenomenon. Aristotle's division of the soul in The Nicomachean Ethics I.13, for instance, has reason on the one hand distinguished from passions, desires, appetites etc. on the other. Hume calls desire a "direct
passion” that “arises from good considered simply” (1978: 438-9). Descartes (1984) and Spinoza (1677) also class it as a passion; further examples could be added. The non-affective, purely function-defined account of desire that might currently seem natural is a historically rather new development (see e.g. James 1997:7, 1998; Rorty 1982).

We might suspect that part of what drives the functional account is nervousness about admitting causal efficacy to affective elements such as feeling. For instance, Kenny (1963:100) argues that when classing desire as a passion, these classical writers were “no doubt thinking of feelings of yearning or longing rather than of the often quite unemotional contexts in which we say “I want”.” And: “Desire, in its most general sense, is not an emotion because it is not sufficiently closely connected with feelings” (ibid.) Kenny distinguishes desire as a sensation from desire as a propensity to act (1963:101), and the account he gives of desire leaves the felt element of it, what he calls the sensation, epiphenomenal both to the intentionality and the behavioural manifestations of desire.42

On a broader philosophical level, the functional account may owe much to the Rylean aversion to ghosts in the machine (Ryle, 1986). The significance for the topic of affect in moral agency is this: insofar as emotion is viewed as suspect, a non-affective account of desire that still held it to be motivating would also suggest that, as long as desire so conceived follows reason, there is no longer a motivational gap in the moral psychology if we remove affect. Shaffer (1983) explicitly uses this kind of approach to give an account of emotion as basically redundant, its causally efficacious parts, according to him, being beliefs and desires. Moreover, he applies Schiffers’s (1976) distinction between reason-following and reason-giving desires43 to argue that for rational agency, only the former kind of desires may be necessary,

42 NB also that Kenny later (1963, Chapter Eleven) explicitly states that a theory of volition is a theory of desire. Presumably, with the affective element in desire relegated to an epiphenomenal role, the worry that affect might sneak into motivation by volition is diminished. On relegation of phenomenal aspects of affect to a merely epiphenomenal role, see also Block’s (1995) distinction between phenomenal and access consciousness, and Warner’s (1980) account of enjoyment.

43 A desire is reason-following if the desirability of its object can be explained (or justified) with reference to some reason that is independent of the desire. So if someone “just wants” a glass of orange juice, their desire for it is reason-providing. If they want it because they believe they should increase their vitamin C intake, their desire is reason-following - though of course there is a potential regress opened as to why getting more vitamin C is desirable. On the vagaries of desire in relation to
and reason-giving desires may be allowed a place in agency insofar as they do not violate some - largely undefined - standards of rationality.

Between Kenny, Ryle and Shaffer, then, we see the outlines of an account of agency where reason reigns supreme, and gets its own tame motivating desires. We might however doubt that desires could be reason-following all the way down, unless - perhaps even if - we simply assimilate desires to something like evaluative beliefs. Also, we might doubt whether the felt nature of desire is in fact entirely epiphenomenal to its behavioural manifestations and in general to its causal efficacy. Even on a dispositional account of desire, desire’s propensity to manifest itself as a felt state might be crucially salient to its motivational impact.\footnote{An epiphenomenal account of affect in desire may also have the undesirable consequence of diminishing the psychological reality of the desires of agents who cannot translate their desires into behaviour: We would not, presumably, want to say that someone who was paralysed had no desires. Attempts to capture the desires of such an agent entirely in the remaining non-affective manifestations of the desire - thoughts, linguistic dispositions, etc., also seem unsatisfactory.}

**Calm passions**

I think it can be conceded to those who worry about emotion that for emotion to be affective in the sense I have argued is crucial to informing and motivating moral agency, it is not necessary for the agent to be in a constant state of psychological upheaval. Nor does such a state of constant upheaval seem desirable. But this is compatible with affect being involved, at a fundamental level, in the dispositions and states that motivate the agent and inform their outlook on the world.

Consider Hume’s argument that what he calls *calm* passions or desires, the ones that have become “settled principles of action” (1978:419) are “often confounded with reason because they produce little emotion” and “are more known by their effects than by the immediate feeling or sensation”. Among these, he includes “certain instincts originally implanted in our natures, such as benevolence and resentment, the love of life” and “the general appetite to good, and aversion to evil” (417). What we call strength of mind, Hume argues, “implies the prevalence of the calm passions above the violent”.

To argue that affect may have a necessary role to play in moral agency is not...
to argue for emotionality in the colloquial sense of excitability and histrionic tendencies. It may be the sensibility and dispositions involved in being emotional that do the main work, not more or less short-lived occurrent felt states. Hume's account also suggests the calm passions have a crucial connection to human flourishing - whose place in morality I will now examine in the next section45.

6) Emotion, morality and flourishing

Hedonic tone and benefits of removing emotion

When discussing possible benefits of removing emotion above, I left for later discussion a fourth type - benefits that relate to the hedonic aspect of emotion. This bears on one of the most notable facts about our emotional experience, which is that it usually - though perhaps significantly not always46 - has *hedonic tone*. Emotion, at least in its episodic and occurrent form, is felt. Feeling emotion is (usually) pleasant, unpleasant or some mixture of the two. So, being emotional creatures, we experience suffering as well as happiness.

Getting rid of emotion - or at least some emotions - might then be thought to save us suffering. While this is not obviously a moral benefit in itself, it is a benefit that morality is plausibly aimed to bring about. Of course, it might be that the positive and the negative experience of emotion will be linked in ways that mean getting rid of the bad will mean getting rid of the good. In some kinds of emotion, notably those involved in attachment to others, the potential positive and negative impact of emotion through feeling may be almost inseparable. Caring about others leaves us open to great hurt as well as great happiness, both through how those

45 NB that a Humean account may still make rather more room for a role for occurrent emotion than an Aristotelian one. While Aristotle gives emotion an important role in moral agency, *pathe* are explicitly subsidiary to their related dispositions in his moral psychology, and his account of virtue is reason-led. For more detailed examinations of the meta-ethical status of *pathe* in Aristotle see Fortenbaugh (1975, 1979), Kosman (1980), Leighton (1982, 1996), Rorty (1984), Cooper (1989), Roberts (1989). On Hume see especially Lind (1990), Baier (1991), but also Blackburn (1998).

46 I shall argue, in Chapter Three, that some emotions may significantly lack definite hedonic tone - the argument is partly based on that in Pugmire (1998). However, I think this can be argued without losing sight of the fact that most feelings seem to fit somewhere on a binary-hedonic continuum, and that this may be crucially important for the motivational impact of emotion.
others behave towards us and through what happens to or is done to them - and the vulnerability increases the closer we get to others (Ben-Ze'ev 1993, Stocker 1997).

More generally, it might also be thought that the capacity for feeling is a precondition of our having emotions at all. (More on this in Chapters Three and Five). And having the capacity for feeling might be not only necessary but also sufficient for triggering actual emotion: if we are capable of being moved, moved we will be. In this case, we might come to the conclusion that in order to be sure of removing emotion, the capacity for feeling would have to go. Whether it would actually be possible to achieve this is another matter: for the moment, the concern is whether it would be desirable 47.

Now, if emotion is a distorting influence, a particularly ruthless stoic may think such losses a price worth paying. Not so much from a pain-pleasure calculus concern with removing suffering: rather, for those who believe that such considerations have little or no place in moral agency, or are even actively undesirable within it, emotion might be thought distorting just for its associations with such calculations. A traditional Kantian could for instance be expected to take this line. For emotion, having hedonic tone, seems conducive to a focus on hedonistic concerns, and hedonistic concerns are often thought to be a distorting influence on agency. And particularly so for moral agency, we might think, since hedonistic motivation need not match moral demands on motivation. Feeling, being typically pleasant or unpleasant, also gives emotion a particularly vivid impact, and this may account for much of the distorting effect on moral perception, cognition and motivation. On the other hand, I shall argue (Chapter Three and onwards), that feeling might also account for a crucial role for emotion in all these three.

Also, the feeling of the emotion seems as much a part of our experiencing the emotion as is the object of the emotion. (The object of an emotion being the thing that the emotion is about or directed towards. Or at least, the thing that the emotion

47 For the purposes of this thesis, it should be sufficient if I can show that removing emotion will not be desirable for moral agency. I shall not deal in depth here with the issue of whether it would be possible to remove emotion, though some of what I have to say about the stoic does bear on the psychological and functional plausibility of agency, especially moral agency, without emotion. (Now usually referred to in the literature as the Mr. Spock issue, after the Star Trek character (cf. Evans 2001, Chapter Two; introduction to Stocker 1996.)
presents as being about - more on objects and their significance in Chapter Three). It may even be that the feeling, and generally the subjective experience, of an emotion commands our attention more than the object of the emotion does - which may seriously affect our capacity for intelligent discrimination. Consider for instance what we could call serial infatuation, where we may suspect the subject is focusing the emotional "high" rather than the object of affection. Such a person might then be ready to discard any particular object of affection if the object or anything about them, from loss of novelty to inconvenient levels of self-assertion, should interfere with the ability to experience the desired feelings. These considerations also point to the addictive potential of emotions.

The impact emotion has through feeling also adds to a previously raised worry about the inappropriateness of emotion. For while we may be able to recognise emotions as inappropriate and so stop ourselves from acting on them, feeling may make both tasks harder. Someone might for instance know, deep down, that they did not have cause to be angry, but the sheer intensity of feeling might not only persist in the face of this knowledge, but also drown it out by force of greater vividness. And even if this drowning out did not occur, the motivational impact of the anger might carry over into action despite awareness that both the anger and the action that followed from it were inappropriate. And a whole range of different intoxicating feelings may be involved in any particular emotion. In anger, for instance, there may not only be the negative feeling of having been wronged, but also the pleasure of anticipating and carrying out retaliation, or pleasure from a consideration of one's own supposed righteousness. More generally, the worry is that people may take pleasure in doing wrong (not necessarily because it is wrong,

48 By what Elster (1999a; Chapter One) might call the reverse mechanism, the person who chases the emotional high might also get fixated on the object: Seeing another person as their one source of true happiness in life, for instance, they might cling like a limpet, often regardless of the effect on the welfare of the supposed object of affection.

49 Cf. Elster 1999b, which also offers some more detailed discussion of analogies and disanalogies between emotional obsession and more literal forms of addiction.

50 Aristotle (The Nicomachean Ethics 1.13) says that in the virtuous person emotion and desire speaks with the same voice as the rational principle. In the non-virtuous cases, it seems, emotion might speak not only with a different voice but louder, in the akratic case sufficiently so to win out over reason in guiding actions. As I have argued above, and will again in Chapters Three and Five, the worry about emotion drowning out reason will be considerably diminished if we allow that reason might go wrong, too.
but regardless), which will reinforce their tendency to keep doing wrong.\textsuperscript{51}

Hedonic tone, then, seems to raise serious worries about emotion. However, as I shall argue at more length later, hedonic tone is also crucial to emotion's role in moral agency, since it anchors our affective engagement with the world. In more everyday language, we care because we can feel. So cases where emotion's hedonic tone leads us astray will be unfortunate local instances of a globally useful feature. And as I shall now argue, hedonic tone is also crucial for the way we as humans are capable of flourishing.

\textbf{Morality and flourishing}

A specific worry about stoicism, then, is that it would undermine links between moral agency and pleasure or happiness, insofar as removing the capacity for emotion might remove these too. Though given some of the worries about hedonic tone attributed to the stoic above, we might suspect they would want to get rid of any sources of hedonistic tendencies, whether emotion-related or not. And at least to the extent that pleasure or happiness is central to human flourishing, stoicism would then also undermine the link between moral agency and human flourishing.

Of course, it \textit{is} possible to take a meta-ethical position of rejecting any place for hedonistic or eudaimonistic concerns in moral agency. Similarly, even if flourishing is granted a place of any importance in morality, it is still possible to operate with a notion of human flourishing that does not see pleasure or happiness or other emotion- or affect-related features as essential to human flourishing. One might, as the Stoics and Kant are thought to do, operate with a notion of human flourishing (at least in its morally relevant form) as purely a flourishing of reason, and possibly also of will\textsuperscript{52}. Concerns with pleasure, happiness etc. would then be allowed only in morally irrelevant areas of agency, if such are thought to exist, and only insofar as the contagion is not likely to spread.\textsuperscript{53}

\textsuperscript{51} Emotion might of course by the same token be argued to reinforce virtue, but since the immediate aim here is to set out what the case against emotion might be, I will not pursue this right away. And a determined pessimist about human nature might think that vice is more likely than virtue to be reinforced.

\textsuperscript{52} Though see Cooper (1996), Sherman (1997), and the notes just below.

\textsuperscript{53} For instance, Nussbaum (1994, Chapter Ten) argues, using Cicero and Seneca as examples, that the Stoics had a "slippery slope" approach to emotion, believing that it would not be possible to
In the absence of emotion, then, we might not feel actual pleasure, but gain in serenity and clarity - though we might query the coherence of offering such states as a tangible benefit of what seems to be a generally *anhedonic* state of being.\(^{54}\) Of course, there may be pleasure in the absence or cessation of suffering. However, the capacity for such pleasure seems to rely fundamentally on a contrast with suffering, and such a contrast could hardly stay vivid enough to register if the contrasting state of suffering never occurred or even figured as a possibility. Knowing that someone you have been worried for is all right is vividly pleasurable by contrast with the preceding anxious state. Knowing they are all right when you have not had cause to worry about them is much less vividly pleasurable, and by extension has less impact.

Of course, we might think that such differentiation by vividness is itself morally questionable (and the prodigal son’s jealous brother would probably agree). At the very least, some degree of critical reflection about it seems warranted. The frame of mind the stoic advocates developing, however, is not just the state of the person who has not recently been concerned in this manner, but the state of the person who simply will not be concerned in this manner in the first place, and therefore fundamentally lacks the relevant kinds of contrast, exaggerated or not.

The starkness of such a view of humanity and morality would not make it particularly appealing, even if it should turn out to be right. And we might think that some place should be made for concerns with pleasure and happiness, both as features of human flourishing and in turn, through their place in human flourishing, in morality. This need not mean taking some staunchly or exclusively hedonistic view, just that we want to rule out the view that such concerns have no place whatsoever, or no important place, in morality. The exact nature of their relation to morality - and more specifically to moral agency - is of course still not fully defined.

---

54 Nussbaum (1994), in discussing the Epicurean and Sceptic arguments that (largely) preceded and in part influenced the Stoics emphasises the notion of "rational uplift" as a benefit of the *ataraxia* (freedom from disturbance) that follows from the extirpation of passions. The Stoics had similar notions: Sherman (1997:112) notes that for the Stoics "apatheia [the absence of passions] is supposed to be *eupatheia*, a wholesome kind of passional state where there is neither total inertia nor the sharp contradictions of passional surges". Interestingly, this state might be close to Hume's notion of "calm passions". As against the aversion to moral risk or risk to happiness implicit in Stoicism, see however the arguments in Nussbaum (1986) and her qualification of the Stoic normative stance of emotion.
Flourishing might for instance be thought of as an end to which moral agency is the means, as an incentive to moral agency, or as a constitutive element of it.

It does not seem possible to rule out purely anhedonic approaches to morality, but their appeal is rather narrow. Perhaps the point is specious, but we might be inclined to think that if morality has no place for such hedonic or eudaimonistic concerns at all, so much the worse for morality.

7) Against moral agency without emotion

To summarise: Against the possible benefits to moral agency of removing emotion, I have argued that although emotion may cause local problems, its radical removal would disable moral agency at a global level. If we were entirely unemotional, we would not be able to recognise ourselves or others as moral agents, and if we could be moral subjects at all, we would be so in a radically different manner than we are now. Similarly, complete emotional disengagement would handicap our capacities for moral perception, to which emotional engagement in the form of participant reactive attitudes forms a crucial background. That we are emotionally engaged with the world through participant reactive attitudes also forms the background to moral motivation, and to our capacity for human flourishing. Some disengaged equivalent of participant reactive attitudes, if such a notion is coherent, would not be able to perform the same function. A capacity for stepping back from our emotional engagements is also a crucial moral skill, but in terms of moral agency, the role of such disengagement is more corrective than constitutive.

The various arguments used for the necessity of emotion to moral agency here can be made independently of each other. However, they are also mutually supportive. For if the capacity for feeling emotion is necessary to human flourishing, which is arguably what morality is for, and necessary for moral perception and moral motivation, that would already make an impressive cumulative case. Add to this my main point, about the consequences of the stoic position for ascription and recognition of moral status, and we have a good overall case for thinking emotion

(1994).
necessary for moral agency. We can then go on to consider, as I will next, whether we may also be able to meet worries about emotion, and what moves we might make towards a fuller understanding of how and why it is that emotion should play such a crucial role in informing, motivating and even constituting moral agency.
Chapter 2: Some worries about emotion in moral agency

1) Introduction

Chapter aims

Chapter One introduced some - rather miscellaneous - examples of worries about the effect of emotion on moral agency. At the time, I put these aside: The first concern was to see whether emotion was necessary for moral agency. Even in the light of a necessary role for emotion, there may still be real concerns about whether emotion also has negative impact on moral agency. The significance of such worries will however obviously be different in this case than if we assume that such worries show emotion to be just plain bad for moral agency. The concern here, then, will be to see how we can best understand and deal with such worries as emotion raises, against a background assumption of the necessity of emotion to moral agency. This examination also throws considerable further light on the nature of emotion and so in the next instance on its role in moral agency.

Of course, if arguments for emotion’s necessary role in moral agency should fail to convince, it will be all the more necessary to have some clear sense of the specific kinds of charges raised against emotion, and of how and whether they might be met. In this chapter, then, I will examine in some more detail what sort of worries we might have about emotion in moral agency.

Three kinds of charges against emotion

I will discuss such worries here under three general headings: Objectivity, Autonomy and Rationality. Emotion has been thought to undermine our capacities for all three. The worry about threats to objectivity is mainly about the distorting effect that emotion may have on perception. As suggested in the previous chapter,

\[1\] In the metaphorical rather than literal sense of perception, though of course emotions may also affect sensory perception, and apperception - things tend to look different to us (larger, more looming) when we are afraid of them, for instance.
there are also worries about how emotion affects our capacity for autonomy, both through making us vulnerable to externals, and through diminishing our self-control. And lastly, there are worries about how emotion affects our capacities for right reasoning, both about how things are (cognitive rationality) and about means and ends (strategic rationality). The worry here may be either that emotion falls outwith the rational, or that it outright violates standards of rationality.

I do not claim that this set of categories is exhaustive. There may be arguments against emotion based on quite different or at least distinct concerns, and I will not attempt a full overview of philosophical worries about emotion in moral agency. Given that such worries can be traced at least to the ancient Greeks, this would be a Herculean task in its own right. However, I do take it that these three headings cover main lines of concern about the impact of emotion on moral agency, and provide quite enough material to be going on with. These three lines of objection can be motivated not just from the philosophical literature, but from folk-psychological assumptions about emotions that anyone might uncover in everyday conversation. On the other hand, as I shall argue, folk psychology, as well as philosophy, also provides some suggestions that the objections can be met.

Some - even most - kinds of concerns about emotion come under more than one of the headings. The three concepts that power the objections to emotion are at least partly, if not always necessarily, interconnected, so that invoking one kind of worry often also entails invoking another. This means, too, that there will be a certain amount of cross-referencing in the discussion that follows - but I will try to keep this as clear and uncluttered as possible.

Looking back at the specific examples of worries about emotion given in chapter 1, some of the worries about emotion raised there might not obviously seem to fall under these headings. Specifically, worries about how emotion affects motivation might not be thought to be covered. I would however argue that these three headings do cover the underlying problems raised by worries about how emotion motivates us. Emotion might be thought to be based on distorted perception, and to cause it: From concerns with objectivity and rationality, then, we might not want to be motivated by emotion. Emotion might also be thought to have negative impact on the strategic rationality of our motivation; to cause behaviour that does not serve to promote our goals. It may be thought to have such counterproductive effects even on the pursuit of goals that the emotion itself has set for us. Moreover, emotion might motivate us in ways we did not choose or want, and which we might feel helpless to control, which would raise worries about autonomy.
There are two main elements to any of these charges against emotion: 1) The descriptive assumptions they make about emotion and its effects; 2) the normative assumptions they make about moral agency. The way these two align with each other will determine how emotion stands with respect to moral agency. So there are two main ways that emotion can be got off the hook: If we can fault the descriptive assumptions that get it in trouble, or if we can fault the normative assumptions that get it in trouble. Since both kinds of assumptions might be due some revision, these are not mutually exclusive moves, and I shall make use of both in what follows.

I begin here by discussing objections to emotion based on concerns with objectivity and autonomy. Since these may pull in opposite directions to each other, as well as being in themselves contentious, the contrast between them motivates the case for emotion as a meta-ethical McGuffin. This is not to say that I dismiss either type of concern altogether: that would be misguided, as well as unnecessary. Both objectivity-concerns and autonomy-concerns also foreshadow rationality-concerns, which I shall discuss last in this chapter. Rationality-concerns are very much to the fore in current debates about what emotion is, which is the topic of the next chapter, so bringing up rationality-concerns at the end here also allows a more natural segue into the question of what difference it makes what emotion is.

2) Objections to emotion from concerns with objectivity

It is a commonplace assumption that emotion is subjective. This is often taken to provide reason to think that emotion raises problems for moral agency. The question, then, is how specifications of this claim bear on issues of moral agency. The worry about moral agency is usually cashed out in terms of claims that objectivity, whatever that may turn out to be, is in some sense necessary for adequate moral agency, and that subjectivity poses a threat to objectivity. Underlying this way of setting out the worry is usually a notion that objectivity and subjectivity, however defined, are such as to conflict. This, I shall argue, is not necessarily right. Also, traditional lines of objection to emotion tend to exaggerate both the subjectivity of emotion and the normative necessity of objectivity.
In the following, I will discuss problems of subjectivity under four headings: *phenomenology, brute feeling, projection,* and *agent-relativity.* In dividing up the topic in this way, I am partly following de Sousa (1987, Chapter Six). He uses rather more subdivisions, though, and does not treat brute feeling as a form of subjectivity. Following Nagel (1979h, 1986), I shall discuss relativity, perspective, indexicality and particularity together under the heading of agent-relativity: the issues are closely enough connected to warrant this. Nagel’s account of agent-relativity arguably also includes phenomenological and projective elements of subjectivity: I have kept these categories distinct here as phenomenology and projection do not by themselves seem to me to raise the kind of issues about *individual* agents that I have grouped together under the agent-relativity heading. I will however discuss briefly, under agent-relativity, how the phenomenological, brute feeling- and projectivity-issues relate to agent-relativity in this more individualist sense.

**Subjectivity as phenomenology**

One main sense of subjective is (following de Sousa, 1987) the *phenomenological* one. The subjective-objective contrast is then set out as a contrast between how things are *in themselves* (objectively) and how they are *for us* (subjectively). Objective truth, so conceived, may be understood as ontologically independent of any subjective awareness: not just observer-independent, but observation-independent.

Subjective and objective in this sense, however, need not conflict. Rather, they may be seen as complementary, the subjective experience being the way in which objective (independent) reality is accessible to us³.

Objective and subjective understood in this way *may* of course also come apart. For instance, in cases of *illusion,* we perceive things that are actually there, but are mistaken as to their properties: The straight stick put into water looks bent. In cases of *hallucinations,* there may be nothing in the world that corresponds to the object in awareness: The pink elephant is not there.

³ Haksar (1991, appendix A) argues that most of the contrasts Nagel offers between subjective and objective (Nagel, 1979h) are of this complementary kind, rather than involving a clash, as Nagel argues they do. Haksar’s argument is, however, not extended to Nagel’s examples from ethics.
Subjectivity understood in this way does not obviously raise problems specifically about emotion. Belief and sense perception, for instance, are subjective, and fallible, in just this kind of way. (de Sousa, 1987: Chapter Six). And this is not usually thought to suggest that emotion and perception are sources of information that we should distrust and discard altogether. Moreover, the assumption that emotion is subjective in this sense suggests that even though fallible, emotional experience does refer to something in the world of which it is the subjective correlate. In that sense, it implicitly places emotion on a par with other modes of perception.

Subjectivity as brute feeling

There is however also a non-representational account of the subjectivity of emotion, in which the what it is like (for us) of subjectivity does not involve intentional awareness in the sense of directed thought-content, but only some kind of brute feeling. On this view, emotion, or at least the felt aspect of it, does not involve any kind of representational content. Such non-representational experiential content may however still have motivational impact on agents, for instance through hedonic tone. Subjectivity, in this sense, might interfere with our capacities to think right and act efficiently the same way that intense physical sensation sometimes does. Or, if we think of it in terms of hedonic tone, the effect of feeling on agency is something like trying to think and plan with a loud boo-hurray chorus in the background.

If this brute feeling has hedonic tone and accompanies states that do have intentional or representational content, such as beliefs, it may also give these states added experiential impact and vividness. This, too, may interfere with rationality, and possibly with autonomy - I will return to this issue briefly under those headings.

---

4 Descartes, in The Meditations, notoriously suggested a global distrust of the deliverances of the senses and any beliefs based on them. But even he makes his way back to accepting information from them, in a qualified way, via his considerations about the non-deceiving God and the capacities of his own finite but God-given reasoning capacities.

5 Wollheim (1999; 6-10) suggests a distinction within the phenomenology of mental states, distinguishing their subjective from their intentional elements, that verges on this, though it is not clear quite how brute his notion of feeling is. And as mentioned in Chapter One, some writers, like Block (1995), distinguish phenomenological content in mental states from representational content, so that phenomenology is understood much as brute feeling is here.
later in this chapter. I will also return to this view in some more detail in the next chapter, where I argue against both the brute view of feeling and a view of emotion as being something like a belief accompanied by brute feeling. We could also add that the concept of emotion I advocate here covers standing, dispositional, calm emotions as well as occurrent felt episodes of emotion. The brute feeling theory seems to apply only to the latter, and I have suggested, in the previous chapter, that the former may be doing a great deal of emotion’s moral work.

As far as normative issues go, the boo-hurray analogy above suggests this account of emotion comes close to a crude version of emotivism about moral thought and motivation. But we might just think that was a further reason to worry: very few are genuinely comfortable to think that moral agency works in such a way. Rather, this would be an account of how moral agency gone wrong would look.

**Subjectivity as projection**

A sense of subjectivity that is both more serious in its implications for moral agency and more specific to emotion is that of *projection*. On the phenomenological account, the subjective is a correlate of the objective - albeit with qualifications about the fallibility of the parallelism. Insofar as the subjective dimension is the one through which we become aware of what is objectively in the world, subjective phenomena, so understood, are even epistemically necessary to us. On the brute feeling account, emotion may have distorting effects on moral agency, but not through misrepresentation. Projection raises in greater earnest the kind of concerns implied in the illusion and hallucination cases: What appears to be in the world is just a shadow cast by the mind.

One worry, then, is that there may be nothing in the world that corresponds to the kinds of features that emotions suggest are there. Instead, there may just be a tendency for emotion-based attributions to be *objectified*, in Mackie’s (1977: Chapter One) sense: the projected feature is treated and experienced as something that is in the world. Witness, for instance, the relatively easy conceptual slide from seeing something as frightening to seeing it as dangerous, or even threatening. “Frightening” describes the object in terms of our reaction to it, but places the
attribution on our side. "Dangerous", on the other hand, applies to the object in itself. And "threatening" implies that there is something in the object that should produce the reaction in us. And this objectifying conceptual slide is not particular to fear: similar examples are easy to come up with for other emotion-types.

Secondly, there is a worry about circularity in the kind of attributions of features emotion makes to the world. Thus we get what de Sousa (1987) calls the Eutypbro-problem of emotion, where we cannot tell which is logically prior: the attribution of a particular trait to the object, or the emotional reaction. For instance, do we love people because we find them lovable, or find them lovable because we love them?

Taken together, these two views moreover suggest that the way we attribute features to objects in the world through emotion may be fundamentally random.

Against this, it could be argued that we do operate with some, however vague, notions of what kind of entities make likely targets for particular emotion-types, and under what descriptions. And the intersubjective consistency of such notions seems greater than we would expect from random projection. Even where we may not find others' emotional reactions appropriate, we may still find them intelligible. Nor are the definitions of the features that warrant particular emotional reactions usually entirely circular: We are usually able to specify at least some features of the typical objects of an emotion without reference to our own reactions to them. I shall have more to say on this in the next chapter, where I discuss formalisation of the intentional objects of emotions.

It may be true that the kind of features that we emotionally respond to only have this kind of salience relative to some sensibility of ours. In that sense, projection is involved. But this is not a sense of projection that need worry us too much, nor is it unique to emotion. Value-perception in general has been argued to be projective on much the same kind of grounds. Also, insofar as such remainder circularity shows that emotional sensibilities are necessary to pick up on certain kinds of features as salient, that would support my argument that emotion has a necessary and irreplaceable part to play.

---

More localised worries about projection

While projection in this global sense need perhaps not worry us overmuch, we might still worry about more localised manifestations of projective tendencies.

For instance, someone who is swamped by anger, even at something trivial, will be inclined to give whatever occasioned their emotional reaction undue importance. And this undue attribution of importance is based in the strength of feeling brought to bear on the object, rather than the object or its significance in itself.

The view of emotion as projective in a bad sense is also reinforced by considering cases of emotional inertia. If emotion does not change with relevant changes in the object that the emotion presents as being about, we would tend to suspect that emotion was not, or at least not entirely, about the object in the first place.

The self-reinforcing effect of emotion may also be thought an example of projection, and of why projection is a worry. For instance, when depressed, we tend to notice depressing things more, and to notice a greater number of depressing things, than in happier emotional states, and may attribute an unrealistic level of predominance to the depressing. We may also find depressing things that are not clearly so in themselves. (Armon-Jones, 1991)

We could also put under the heading of projection the tendency for emotion to spread itself onto objects other than those that it originally was, or seemed to be,
about. For example, after an argument with a significant other, people are often more inclined to be angry at third parties, or in a general mood of more or less unspecific irritability. Mood also shows a tendency to spill over in the opposite direction, that is, to go from being more or less unspecific to fixing on particular objects - as when general irritability is funneled into anger with a specific person or issue (Wessman 1979; Armon-Jones, 1991). So other affective states may be prone to the same worries as emotion here.

Emotional projection may also be crucially involved in wishful thinking. And emotion typically features in projection in the psychoanalytic sense: the attribution to others of mental states that are actually one's own. In all these cases, the mind casts a shadow onto some object external to it. In wishful thinking, the shadow cast by how one would rather things were is taken to be how things actually are. Or in the psychoanalytic case, one's preferences about our own and other people's mental states casts the shadow. We might wish, for instance, that others feel the same as us, whether straightforwardly, as when we want them to find the same things interesting, pleasant, or offensive, or less straightforwardly, as when we wish for them to reciprocate our feelings. Or we might wish that we were not feeling what we are feeling, and so attribute the emotion to someone else: "It's not me that is feeling aggressive here: It's the other person".

In all these cases, it seems that what we find in the world as emotion presents it to us is there only because emotion put it there, or at least was involved in putting it there. Or if not emotion, then most likely some other form of affect. Though it on the depressive realism thesis.

In fact, desire is often seen as the crucial distorting agent, and emotion as guilty only by association. Mele (1997, 2000/1), in discussing the role of emotion in self-deception, raises the possibility of emotion-only confirmation bias, though he does not seem overly convinced of its plausibility, and is notably short on possible examples. Moods, by contrast, seem to offer plausible examples of distortion without desire-involvement: Even if we accept a case for "depressive realism", we might still indict mania and anxiety. And perhaps contentment might be indicted, by extension of the depressive realism thesis.

The rather various ways in which desire might feature in relation to other affective types complicate the picture here. For instance, specific emotion-types might be thought to entail characteristic desires, as anger is thought to involve a desire for retaliation. Moods may also entail desires, albeit often rather unclear ones: diffuse urges to escape or seek safety in the case of anxiety, diffuse aggressive urges in irritability. Both emotions and moods might follow from the satisfaction or frustration of desire: Gordon (1987) and Wollheim (1999) both argue, albeit in rather different ways, for this as the originating condition of emotions as such. Or finally, particular emotion-types, such as romantic love, might be thought desirable ones to feel, whether because feeling them is in
should be noted that there may also be cold (unmotivated, non-affective) sources of such reinforcement of emotion. Mele (1997) cites confirmation bias, our tendency to look for and notice evidence for rather than against any hypothesis in question, and a bias toward vivid information. Once a particular emotion is present, we might expect that both of these would tend towards reinforcing it.

If we accept projection as being unproblematic per se, we can no longer draw a contrast between objective and subjective in terms of an in-the-world/for-us contrast, since we will have accepted that there is nothing as it were in the world corresponding to our perceptions of wrong or right. But as already argued, this need not mean giving up on criteria of validity altogether, or just letting anything pass. And the kind of cases suggested just above seem prima facie candidates for what we would want such standards to rule out as wrong or inappropriate.

This of course still leaves legitimate worries about how we set criteria of validity or appropriateness for projections, and on what basis. Appeals to human nature, the general point of view, detached observers or universalisability as bases for standards all have their own notorious problems, some of which will also turn up in the next section, on subjectivity as agent-relativity. But insofar as all these ways of setting up standards of right moral response while allowing for projection in moral perception have venerable traditions, too, allowing for the possibility of projection being involved in emotion would not necessarily be that damaging, either.

The sense in which emotion involves bad subjectivity, then, may not be that it is projective, but that the projection involved fails to fulfil relevant criteria of validity. That such cases may by rationalisation or other means present as valid, and not in all cases be detectable, is a real worry, but on a local rather than global scale. And these cases are not only anomalous to, but also parasitic on, cases of more appropriate emotion. Some more will be said about criteria of appropriateness for emotions under the heading of concerns with rationality, below. I shall also have

---

Itself pleasant, or for more “ideological” reasons. (See also de Sousa, 1980, on self-deception and emotion).

This last point also suggests an incentive to deceive oneself into thinking that one has the emotion in question. Hence, perhaps, phenomena like adolescent “girly” crushes, which might be seen as motivated self-ascriptions of being in love, specification of the object of affection often seemingly tagged on as an afterthought.
some more to say about how we may deal with emotions that reflect those who feel 
them more than what they are “about”, in the next two chapters. First, though, a final 
sense of subjective.

**Subjectivity as agent-relativity**

There is a common use of subjective in ordinary language to mean 
something like particular to the individual. Thus, to say that someone’s outlook is 
‘just subjective’ is to indicate that others may not see things the same way. It may 
also further suggest that others are not epistemically, morally, or otherwise, required 
to see things the same way. Basically, the worry seems to be that differences 
between agents raise concerns for moral agency. Insofar as several takes on this kind 
of subjectivity are more or less related, they will be discussed together here under 
the heading of agent-relativity.

The kind of agent-relativity I will discuss here is what we could call the 
*individualist* version. Phenomenology, brute feeling and projection are all agent-
relative in capturing how things are *for agents*. But agent-relative subjectivity in 
those senses could at least theoretically be the same for all agents. The concern here 
is with *particular subjectivity*.

This kind of concern has a number of distinct if not always easily separable 
sources. For instance, the problem of subjectivity may be seen as one of *relativity*. 
Emotion is often thought especially liable to relativisation to the particular features 
of the individual agent, their situation, and their interests. Insofar as agents can be 
assumed to vary considerably with respect to all these, their emotions can be 
expected to vary accordingly. Insofar as such differences in emotion tend to be 
further reflected in different judgements of value, action-tendencies, and so on, 
emotional subjectivity brings in all the familiar meta-ethical problems of moral 
relativity.

And this also points to worries about *indexical, particular and idiosyncratic* 
features of agents. The emotional reactions of individual agents are coloured by their 
particular histories and personalities, their roles and situations and their positions 
relative to other agents, and at any point in time and space, by what is most vividly
salient to them at that point (Ben-Ze’ev 1987, Blum 1994, Stocker 1997). And each
agent occupies their own subjective world, which will differ from the subjective
worlds of other agents, as well as from the objective world. (Nagel, 1986). The
subjective world may be thought of in terms of (situated) dramatic (de Sousa, 1987)
or narrative expectations (Marshall 1980; Hepburn 1984; Goldie 2000a; this notion
is also a consistent theme of Nussbaum 1986, 1990). With regard to these
expectations, different agents will at least to some degree also have different
emotional idiolects (de Sousa, 1987), of which more in the next chapter.

Another way of understanding subjectivity in this sense is to view it as a
matter of perspective or point of view, which may itself be taken to involve
relativity, particularity and indexicality. Nagel (1986) argues that just as each
perspective brings with it a particular outlook, it also brings with it particular blind
spots. Each person’s perspective is qualitatively and numerically distinct from those
of others. Experientially, epistemically, and motivationally, it makes a difference
that I occupy my particular perspective and not another. Nagel’s argument in “What
is it like to be a bat?” (1979g), also suggests that for any particular perspective, what
it is like to occupy that perspective may be more or less inaccessible from outside it.

And this not only limits the degree to which emotional reactions are shared
by and common between agents, but also the degree to which they are
intersubjectively intelligible. The particular perspective, with its associated blind
spots, may also blinker the moral perceptions of agents. And the intimate connection
each of us has to his or her own perspective, and lacks to the perspectives of others,
may also be expected to affect motivation. This seems particularly significant since a
recurring worry, for all the senses of agent-relativity distinguished here, is that it is
partial, in the partisan sense, as opposed to impartial, as well as being partial as
opposed to comprehensive.

Insofar as emotion is understood to be agent-relative, and insofar as agent-
relativity may be a handicap for moral agency, this particular reading of the
subjectivity-charge raises real problems for a vindication of emotion in moral
agency. The more so in that concerns about agent-relativity may combine with
concerns about projection. Even those, like Hume, who would not be unduly
troubled by projectivism about moral values typically draw the line at allowing, let alone validating, projection from idiosyncratic, relative perspectives. And of course, if we combine some account of moral value as being in the world with an account of emotion as inherently projective and relative to particular agents, emotion may look particularly bad. Both as a meta-ethical account and as an account of emotion, though, that reading would have considerable problems of its own.

Some mitigating points to worries about agent-relativity

The picture sketched above also belies the extent to which we are capable of overcoming the barriers set up by our particular natures. We are able to, and in fact we often do, transcend our own particular perspectives. We may take into account, understand and even attempt to share the points of view of others. And we may have emotions on behalf of others, as well as on our own behalf. On the point of motivation, as well as perception, Strawson’s intersubjective account springs to mind here: most of us are capable of, and do make, demands for others and of ourselves.

Another common way to approach worries about particular subjectivity, though, is to advocate not these moves from one’s own particular points of view to those of others, but a move from a particular to a general point of view. But this move, too, is one we are emotionally capable of. And this ability is in fact something we crucially draw on for moral information. If I want to examine the moral character of my own actions, for instance, I may try to imagine them as they would look when viewed from outside but not by anyone in particular, and see how I feel about them then.

Taken to its logical conclusion, the move away from particular perspectives would end up in Nagel’s (1986) view from nowhere, a centreless transcendent perspective on the world. However, Nagel argues, the view from nowhere does not

---

10 Williams (1973c) contrasts a “Humean leap” from self to other from a “Kantian leap” from particular and affective to universal and rational and argues that both for moral psychology and for moral theory, it is the Humean move that makes the difference.  

11 The standards by which I judge my actions may still be more or less relative: In viewing myself from the outside and from the point of view of no particular person, I may still be judging by moral standards that are not shared by all communities (see e.g. Williams 1983 ch.5, Gibbard 1990 chapter 3, on “communities of judgment”). Insofar as lack of universal moral standards is a problem,
encompass the particular perspective that was the starting-point, or indeed any particular internal perspective. So even the view from nowhere will have blind spots. Nagel, on this basis, argues the need to find some way of combining the insights of both the transcendent general and the finite particular perspectives.

And here is a point of contention for normative assumptions. For while some particularly hard-line impartialists, such as Kagan (1989), and arguably early Nagel (1970) have insisted that only the impartial, transcendent perspective can reveal moral truth to us, and that only those things have value that are valued from that perspective, others have denied this. And many who see a value in adopting the transcendental perspective reject the view that it is the only one we should take into account morally. Nor need we think that any particular individual’s concerns and projects have no value because they appear insignificant from some more depersonalised or impartial perspective.

As against the line taken by Kagan, for instance, Blum (1980a, 1994), McNaughton (1988), Dancy (1993), Crisp and Korsgaard (1996) all argue for the indefeasibility of agent-relative values - Dancy and Crisp specifically target Kagan’s account. The battle-lines often tend to be drawn in this way, with deontologists, virtue theorists and particularists siding against utilitarians. We might however think that the utilitarian could have reasons of his own to make room for agent-relative values. For the basic utility-bearing unit is the individual agent, and we might reasonably expect that their agent-relative features would have some bearing on what would count as utility for them.

None of this is to argue for accepting any and all agent-relative concerns as morally valid or permissible: Clearly, we would want some standards to apply here too.

While we can motivate some real worries about intersubjective dimensions of moral agency being under threat from the structure of particular subjectivity, then, it is not clear that the way to deal with this is to try to eradicate particular subjectivity.

though, it is not necessarily a problem caused by emotion.
Also, a reasonably internally coherent and stably structured particular subjectivity itself seems to have a necessary role to play in moral agency. It may be important, for instance, for autonomy, and for what Wollheim (1984) calls a *sense of self*: The awareness which a person, any person, has of himself, and which allows him to think of himself as an ongoing person: influenced by his past, living in the present, and concerned for his future. Both a capacity for autonomy and a sense of self would seem crucially important for dealing well with the intersubjective balancing act Strawson thinks a normal moral agent will be performing. A big worry about agent-relativity is that it inclines us towards solipsism, each agent within their own little subjective world. But it should not be necessary to lose one's self, or one’s own perspective, to gain a world with others in it.

Awareness of agent-relative features, one’s own and other people’s, also seems crucial for managing moral agency. For to be able to deal well with people, ethically, one needs some sense of who they are. For many kinds of interaction, only a very thin sense of what people are like may be necessary. In these cases, one may get by morally on very general considerations about respect for persons, or about how one may or may not treat sentient beings. The closer the interaction, though, the thicker the sense of their natures required. This also has implications for how one deals with oneself, as moral agent and as moral subject, as Vetlesen’s qualification of Strawson, cited in the previous chapter, indicates. For from others, and the need for a sense of them, escape is possible, if only temporarily. From the self there may be no real reprieve – or at least not as long as one is alive and basically capable enough *qua* agent to count as a moral agent.

In managing all these aspect of moral agency, emotion provides crucial information and motivation. But management of emotion is also crucial. Since emotions can go wrong, the ways they make us think, be motivated and act may require critical scrutiny. But this need not show emotion to be somehow uniquely dubious, or even that it is more likely than not to be wrong.

---

12 On reductivism about persons in conceptions of moral agency, see also Chappell (1997).
Subjectivity and moral agency

Subjectivity, in its various senses, is a concern for moral agency mainly as a potential epistemic handicap. The worry for moral agency is on two distinct lines. First, there is a concern that non-moral information moral judgements build on may be distorted by the presence of emotion. Emotion has been held to have a generally detrimental effect on our capacities for correct perception, and thereby to undermine basic capacities for efficient agency, as well as those more specifically concerned with moral agency.

Secondly, there is a worry that our moral standards and judgments themselves may be distorted by emotions. Exactly how this claim is cashed out will depend crucially on what account is given of the ontology of value, and what sense of subjective is applied to emotion.

But with the possible exception of the brute feeling view, of which more in the next chapter, the descriptive accounts above do not seem to warrant such strong conclusions. While they give reasons to be concerned about emotions, they do not provide a basis for a strong case against them. On the normative side, the case against emotion from subjectivity is also limited by the consideration that emotion also has a crucial role in what needs to be done to sort out the problems.

The fact that we do speak of emotions, both in philosophy and in folk psychology, as warranted or not gives further hope that emotion may be subject to and responsive to criteria for validity or at least appropriateness. We might also hope that this leaves room for some resolution of issues around agent-relativity. I will return to the notion of criteria in this chapter in discussing rationality-issues around emotion, and in the next chapter.

3) Objections to emotion from concerns with autonomy

Introducing worries about autonomy in the previous chapter, I divided these broadly into external and internal failures of autonomy. Insofar as emotion leaves us hostages to fortune, and with hostages to fortune, autonomy is threatened from
outside. Emotion may however also threaten autonomy from within, damaging and 
undermining the agent’s capacities for autonomy.

These two kinds of worry about autonomy may moreover feed into each 
other. Loss of capacity for autonomous agency from within may lead to greater 
vulnerability and susceptibility to externals. Conversely, the effects of vulnerability 
to externals may undermine the capacities for autonomy in the agent him/herself. Following Kant, I will tend to use heteronomy as a shorthand term for the more 
clunky “failure of autonomy” in what follows.

As concepts of autonomy vary a good deal, I will begin by briefly setting out 
some main contenders, and some important related concepts, such as integrity. I will 
then look at how emotion stands with respect to these concepts, and finally briefly 
raise some worries about the value of autonomy itself.

**Concepts of autonomy and concepts relevant for autonomy**

The importance of any charges against emotion from concerns for autonomy 
will of course depend on the concept of autonomy employed. A large part of the 
literature on autonomy is taken up with issues of definition, and concepts of 
autonomy may vary considerably between authors. Comprehensive discussion of the 
concept falls outside the scope of this thesis, but some scene-setting will be 
necessary for reference in later discussion.

Given pride of place in moral agency by Kant, concepts of autonomy still 
feature heavily - often implicitly rather than explicitly - in both philosophical and 
non-philosophical thought about moral agency. Extensionally at least, concepts of 
autonomy and of moral agency often overlap to some degree: autonomy often 
figures in constituting and defining moral agency as well as in evaluating it. On a 
Kantian picture, for instance, autonomy is defined in terms of the possession of 
reason and free will13. Through those, autonomy is the capacity for rational self-
determination (Selbstgesetzgebung) - which leaves it near enough co-extensional 
and co-intensional with the Kantian concept of moral agency.

---

13 There are of course notorious and stubborn problems about free will, which in turn make 
using it as a central definiens of autonomy difficult. Nor need ascriptions of free will, moral agency 
and responsibility coincide exactly - see e.g. Watson (1987) and Greenspan (1987), both in Schoeman
Across the varying definitions of autonomy, this notion of being self-determining occupies a crucial place, even though its exact meaning varies and is often crucially vague and inexplicit. One way to try to capture this connection between self-determination and autonomous moral agency is in terms of Strawson’s distinction between participant reactive and objective attitudes. Taking the participant reactive attitude towards someone, we see them as responsible or at least answerable for their actions, attitudes and so on. Taking the objective attitude towards someone, we class them more as a thing than as a person: in the Kantian sense of the term, we objectify them. Things cannot be self-determining. What they are like is not up to them. What they do is more like an event than an action. For all that we may feel a range of emotions towards them, we will not feel those reactions that depend on the notion that this is another person freely and deliberately doing this towards me. A person’s actions and behaviours obey the person’s own laws (which is the literal Greek meaning of autonomous), but an object is acted on by laws whose sources are other (Greek heteros) than itself.

For instance, we are typically said to be autonomous insofar as we are active with respect to our attitudes, behaviours and so on, rather than passive. Or autonomy may be thought to be essentially about choice: we are heteronomous with respect to something insofar as we are not exercising, or capable of exercising, choice with respect to it. Or again, autonomy may be understood to be about being oneself the (active) source of one’s attitudes, behaviours, and so on. (de Sousa, 1987, gives this as the existentialist sense of autonomy). Though again others (e.g. Elster, 1983) think that one can be autonomous in some attitude insofar as one has chosen it, without the attitude having to be original to oneself. Or one might insist that the choices involved in autonomy should be ones the agent identifies with (Frankfurt, 1987), as opposed to ones they merely find themselves making - which brings back the point about passivity.

Or we might emphasize issues of what degree of control the agent is capable of over the area in question. Thus Kant, in attempting to limit ascriptions of moral (1987). Fortunately, dealing with these issues in depth falls outside this thesis.
agency to areas in which he felt we could be held responsible, denied that the emotions and more generally the inclinations could be part of our moral agency.

For non-Kantians, though, moral agency may also extend to aspects of ourselves for which we are not usually thought responsible - hence concepts of *moral luck* (Nagel 1979b, Williams 1981b). For aspects of ourselves for which we are not strictly held responsible, we may still be thought answerable in some less stringent sense. Sabini and Silver (1987) suggest this is the case for character and emotion, with respect to which we are legitimate targets for judgement. The judgement, though, is rather more like that involved in aesthetic evaluation than that normally applied to more straightforwardly *moral* aspects of agency (but see also Morris, 1987, on "non-moral guilt"). Character would here be understood to include emotion-related and temperamental traits, all of what Murdoch (1985:102) argues Kant left out in attempting to "segregate the messy warm empirical psyche from the clean operations of the reason". Such critical accounts can still allow that we are in some respects autonomous, and that this an important, even crucially important capacity of moral agents, but autonomy and moral agency will not coincide as neatly here as they do in the Kantian approach. I will return to some related points in discussing the value of autonomy at the end of this section.

Given the crucial link between notions of moral agency and of autonomy, we might also expect different descriptive and normative notions of agency to make a difference here. For instance, some hold that to be an autonomous agent it is necessary to exhibit a high degree of internal coherence as an agent, with respect to the mutual compatibility of one's different attitudes, goals and practices. This view could for instance be reached by extension of the notion of autonomy as self-legislation. Or we might think a capacity for forming efficient and considered higher-order constitutive of autonomy (Frankfbrt, 1971, 1987) or at least instrumentally necessary for it. On the other hand, more existentialist approaches may instead put a preference on the gratuitous, the spontaneous, the immediately sincere, as being crucial to autonomy, and not insist on intrapersonal consistency.

*Integrity* is often thought crucial to autonomy, though this assumption may be given a different slant depending on what it is thought desirable that we be
autonomous with respect to. For someone with a more individualist account, we might need integrity to protect our autonomy from conformity. Others might not think conformity per se a problem, but still be concerned to protect autonomy, possibly through integrity, from mere yielding to social pressure. A Kantian would for instance presumably not be concerned with conformity of opinion per se, but with the validity of the process by which opinions had been formed. In fact, the Kantian emphasis on universalisability, in the first formulation of the Categorical imperative, would suggest conformity would be a desired outcome. Which it not to say that conformity was the end sought. Others, of a more existentialist bent, might insist on individuality per se as a marker of autonomy. A Kantian would presumably not be concerned with conformity of opinion per se, but with the validity of the process by which opinions had been formed. In fact, the Kantian emphasis on universalisability, in the first formulation of the Categorical imperative, would suggest conformity would be a desired outcome. Which it not to say that conformity was the end sought. Others, of a more existentialist bent, might insist on individuality per se as a marker of autonomy.14

Autonomy is also often associated with independence, though what it is thought important to be independent with respect to varies. In some cases, claims for the importance of independence may not amount to much more than ringing the changes on the notion of the agent being the source of their own features (such as their emotional reactions), or at least exercising active choice with respect to them.

On the other hand, independence from externals seems to be a naturally central concern for any concept of autonomy. So the important thing might be integrity, to avoid the agent’s being influenced unduly and without careful consideration of merit by the views of others. Or it might be a capacity to avoid or at least rise above attachments that might be thought to exert undue influence. This may, conversely, give rise to worries that the autonomous person may be unduly disinclined to become engaged in interpersonal relations or relationships of any kind (Onora O’Neill, Gifford lectures, Edinburgh 2001).

Or the independence intended may be the kind of blithe ability to rise above circumstance envisioned by the Stoics: All sorts of misfortunes may befall the agent, but insofar as they are autonomous, they will not be affected in themselves. Even if they find themselves like Job after disaster struck - disfigured, diseased, bereaved, impoverished, scrabbling on the dustheap, surrounded by friends who suggest these misfortunes must somehow be all their own fault - they will not let it get to them,

14 For a broader critical discussion of the connection between autonomy and sense of self, see e.g. Young (1980).
but will determine their own attitudes by some inner standard unaffected by externals.  

On the point of independence, it is worth pointing out that autonomy generally has this relativised aspect: We are autonomous (or not), within ourselves, with respect to our attitudes, beliefs, values, behaviours, or preferences. In relation to the outside world, we are autonomous - or not - relative to other agents, to the vagaries of our physical environment, to surrounding social and historical contexts. As indicated at the outset, there may be threats to autonomy from both inside and outside. The force of general opinion may threaten autonomy from without, but the flotsam of one’s own nature may threaten it from within.

Insofar as any or all of the above considerations are relevant to issues of autonomy and emotion, they will be drawn on in what follows, without attempts to come down conclusively on the side of any particular account of autonomy. The question now however is how all this bears on issues of emotion’s place in moral agency.

**Emotion and autonomy**

_Are there emotion-specific threats to autonomy?_

If emotion poses a threat to autonomy, then, it may also by extension pose a threat to capacities for adequate moral agency - or moral agency at all. So in what ways could emotion specifically be thought to present a threat to autonomy?

Well, for one, there are always areas, for any merely human agent, in which we do not exercise choice or control. Some of those areas are fairly clearly external to us - our environment, other agents. Of course, our powerlessness here may often be a matter of degrees rather than absolutes. Moreover, we may impose deliberate boundaries on the exercise of such power as we do have, for instance in the cause of respecting the autonomy of others: so insofar as we aim to be moral agents, that may in itself limit our outward capacity for exercise of power. Other areas where we lack control or choice may however fall within ourselves. We seem to have no more

---

15 For approving examples of how people might rise above even the most extreme circumstances and not succumb to emotion, see also Seneca’s “On anger” (in Seneca, 1995).
choice, at least initially, with respect to our own natures, abilities and range of capacities than we do with respect to externals. And the boundaries may not be clear here: externals may influence and shape aspects of ourselves that are themselves not, or not fully, under our control, and the distinction between outside and inside may in some areas be significantly elusive.

Coming to understand, and coming to practical and theoretical terms with, the boundaries within which we may genuinely be said to exercise agency is a crucial skill for a moral agent to master.

Emotion, however, tends to give impact on us to things outside our control. In fact, most of the things that matter to us, emotionally, may turn out to be things outside our control. And some of the reasons why this matters are specific to emotion. If emotion merely registered what went on in the world, it would not be so bad. However, emotion has greater impact than that, for through emotion we do not just notice what goes on, we feel it in ourselves. In some cases, the emotions with which we register things or events outside our control may be so overwhelming that our capacities for any kind of agency are debilitated, as sometimes happens in cases of intense grief, depression or fear.

Even where agency is not altogether crippled by emotion, which is after all the majority of cases, emotion may still have powerful distorting effects that go directly to issues of autonomy. We may be swayed cognitively, motivationally and behaviourally by the pain and pleasure of our emotions about externals in ways which are not ones we should have chosen independently of that pain or pleasure. The ways emotions move us may go directly against what our considered opinions suggest - and sometimes, emotion carries over into action seemingly against our will.

The impact that emotion gives to externals also leaves us with a temptation to try to manipulate outcomes, a temptation which may have undesirable effects on moral competence. This is of course particularly a problem from a Kantian deontological and anti-consequentialist point of view. Under the influence of emotion, we may do, not what we think we ought to, but what we think will produce some spuriously appealing outcome.
This seems all the more likely in that short-term interests are usually more vivid to us than long-term interests. Long-term interests seem likely, in a reasonably self-controlled agent, to be ones we can settle on in ways that support autonomy. However, a stray emotion of the moment, setting up its own short-term context of interest, may distract from this long-term path, temporarily or permanently. If we feel fear at the moment, our main immediate concern is to avoid the object of fear, so as to diminish the unpleasant feeling. Or the short-term happiness brought by some pleasure that is not in our long-term interest may lead to the repeated short-term pursuit of that pleasure, as in addictive behaviour (Elster, 1999b). And the cumulative effect of this is to undermine the original long-term goals, which would also make emotion problematic on most consequentialist approaches.

The combination of awareness of limitations on one’s own power with the impact emotion gives things outside one’s control may also distort autonomy with a different direction of fit. Rather than try to suit the world to one’s own preferences, one may try to suit one’s preferences to the world. We might, for instance, stop wanting things because we do not believe them attainable (“sour grapes”). Such adaptive preference (Elster, 1983) may be viewed as inauthentic, as well as irrational, and as undermining of autonomous preference formation.

**Emotion and intersubjective heteronomy**

There is also what I shall call *intersubjective heteronomy* in emotion. This takes several distinct forms. Concerns about the actions, attitudes and opinions of others may distort our capacities for autonomous agency. Intersubjective heteronomy may also come about through the interpersonally contagious nature of attitudes and feelings (see e.g. Goldie, 1999). Both kinds of intersubjective heteronomy may for instance be involved in giving in to persuasion. To take a trivial example, heteronomously going for a social drink despite intending to have a sober and early night might not “happen to” someone just because company and drink are an appealing combination, or because of more or less subtle social pressures brought to bear. The shared feeling of expansive, expectant happiness in the group may also be catching. And less trivial problems about asserting one’s autonomy with
significant others, in romantic or family contexts, for instance, may involve such challenges too (see e.g. Friedman, 1998).

Wollheim (1999, lecture 3) suggests that at least some of our emotions may be even more deeply intersubjectively heteronomous than this. Guilt and shame, for instance, are learnt in circumstances where the agent is under pressure to conform to standards not his own: Wollheim calls this mere heteronomy. He distinguishes this from radical heteronomy, where the attitudes and feelings others have towards us affect us even where we ourselves no longer have the attitudes that would warrant being so affected. We may feel shame, for instance, not because we think we have anything to be ashamed of, but because we are aware that we appear to others to have something to be ashamed of.

While Wollheim is specifically talking about moral emotions (or as he has it, "the so-called moral emotions"), these points extend, mutatis mutandis, to emotions in general. For we learn to identify emotions, their appropriate occasions, and the desirability or otherwise of particular emotional reactions, in social contexts that are not of our choosing or within our control. And even where we might come to diverge from this original context in our emotional reactions and standards later in life, something of the original imprinting often lingers. And where this is so, reactions appropriate to the imprinting can be triggered by the attitudes of others. Someone who has on the whole come to terms with being homosexual may for instance still find it hard to maintain a positive self-image in the face of homophobic attitudes.

The original imprinting may also linger in the form of willed opposition to it, which would also be heteronomous. When someone adopts a flamboyantly outrageous lifestyle in reaction to a repressive upbringing, or a rigidly neo-conservative lifestyle in reaction to their baby boomer counter-cultural parents, we would hardly think of them as being autonomous. Compared to someone who just passively and unthinkingly conforms, such agents are at least making a bid for autonomy - but the difference between them and a genuinely autonomous agent is rather like the difference between an adolescent and an adult. Granting, of course,
that the notion of a truly autonomous adult agent may be an idealisation with few if any real instances.

Altogether, then, intersubjective heteronomy opens possibilities for agents to manipulate each other through emotions, as well as for undermining autonomy in less overtly intentional ways. And our responsiveness to such manipulation suggests an other-directedness in our emotional natures that could be seen as a fundamental flaw in our capacity for internal autonomy. Though as I shall argue in Chapter Four, the problem is with the other-directedness rather than with emotion as such. And the other-directedness may even have crucial uses of its own, to avert the equally worrying, and equally human, tendency towards emotional solipsism16.

Emotion, control and choice

Apart from these worries about external threats to autonomy, emotions are themselves often said to be outside our control. We do not normally decide what we feel, what about, and how strongly. Or at least, we do not do so in any simple and straightforward way. We cannot usually produce desirable emotions at will, nor can we usually prevent having undesirable emotions. And once present, emotions may not be capable of much adjustment in degree, in kind, or with respect to choice of objects, by will or reason. Or again, at least not in any immediate or straightforward way. Our emotions may appear to come upon us and remain with us regardless of, or even against, will and reason. Insofar as we are capable of controlling them, which is usually indirectly, by direction of attention, persistent arguments, or through the control of behaviour, our control may come at high cost in time and energy. Moreover, it may fail, as in akratic cases, where the motivational impact of emotion means it carries over into action even against the opposition of will and reason.17

These features of emotion may themselves be instances of the impact of externals, insofar as emotion is understood as a reaction to its object, and salient features of the object are understood to determine the nature of the emotion. This is

16 On emotions as communications, see e.g. Oatley (1996), Griffiths (1997).
17 NB that such loss of control may be interpreted as a form of "disclaimed action", protected and possibly constituted by social conventions about the involuntary nature of (strong) emotion and its associated behaviour. Cf. Griffiths (1997, ch.6). Some constructivist readings of emotion (Sartre 1948, Solomon 1993), think this implies conscious but disclaimed bad faith. I will discuss this again
also an important sense in which we might be said to be passive with respect to emotion. On the other hand, a similar kind of passivity may apply in the case of belief, where it is not thought to be problematic. In fact, the converse idea - that belief might be a matter of choice, and motivated choice at that - is the one more commonly viewed as problematic. I shall return to this issue in the next chapter, in discussing judgementalist theories of emotion.

Also, if emotion is heteronomous because it is determined by the nature of the object, and not by the person who feels it, it seems that autonomy, by contrast, involves something like projection. But this projection, in turn, must be active, not passive. For by itself, the view that emotion reflects the person who feels it instead of what it is felt about merely leads back to the question of whether we possess much autonomy in relation to our own natures. If becoming attached to others, for instance, is something over which we exercise little control, and which is triggered by features of those others, and of ourselves, over which we exercise little control, it will still be heteronomous - and overdeterminedly heteronomous, at that.

Of course, insofar as emotion does reflect aspects of our natures with respect to which we are autonomous, it could reflect and even aid autonomy, but the suspicion may linger that emotion's natural tendency is not in this direction, and that its reflecting creditable rather than discreditable aspects of our natures is too much a matter of luck. This worry is reinforced by the consideration that formative emotional experiences shape character traits in later life, and that such formative experiences are not under our control. Which also suggests that character-formation is vulnerable to moral luck, and by extension throws suspicion on emotions determined by that character.

On the other hand, the worry is sometimes that the emotion is too much determined by a character whose formation was not autonomous, but that emotion is out of character. The emotion may in such cases be experienced by the agent something in the manner of an alien takeover or possession: he cannot understand why he is feeling this particular emotion, in this degree, towards this particular in the next chapter.

18 The related possibility that emotion itself is overdetermined, and its implications, is more fully discussed in Marshall (1980).
object, when neither the emotion nor the choice of object may be in keeping with his general outlook or even his general emotional profile. In these cases, the agent may experience himself as internally divided as agent. And in fact, the agent may be so divided. But such internal division is *prima facie* undesirable. And emotion may be thought to be causing it. At the very least, emotion may be a manifestation of such internal division, and one likely to reinforce the division, at that. And so emotion again comes out looking dubious.

We might still think, as I shall argue here, that whether or not these tendencies manifest themselves as actual heteronomy may depend on features of the agent other than whether they have emotions or not - character traits, for instance. More generally, it also seems worth pointing out that tendency is not necessity. And, I shall argue, in the later chapters, emotions may have a crucial role in increasing autonomy, through the effect that emotional readjustment has on agency.

**Some worries about the value of autonomy**

**Doubts about radical autonomy**

On the face of it, then, there are important ways in which emotion raises problems for autonomy. But based on exactly the kind of examples used against emotion, above, we might also have some worries about the value of autonomy, and about what kind of concept of autonomy we ought to promote for moral agents.\(^{19}\)

One such converse worry about autonomy and emotion is this: Complete choice with respect to our emotions may be at the least utopian, and possibly incoherent, as an ideal. Radical choice in this respect may be a chimera both in the sense of being unreal and in the sense of being monstrous. This worry applies mainly to the more existentialist versions of the account. Those that operate with some, however minimal, notion of human nature, may be able to work that into their account of autonomy as a given from which autonomous choices can proceed and against which they can be measured. Consider, for instance, Kant’s notion of the

---

\(^{19}\) For more general discussions of the value of autonomy, see Young (1982) and also Onora O’Neill’s recent (2001/2) Reith and Gifford lectures on autonomy and trust.
categorical imperative as a constitutive principle of rationality, and rationality as a constitutive element of personal agency.

This latter kind of account gives some hope about the chances of reducing the heteronomy-charges against emotions, as long as we are allowed to bring some more of the messy warm empirical psyche into the picture, as well as reason.

The way to this account may be through addressing a related kind of worry about emotion and autonomy, suggested above: that there is some tension between notions of autonomy, which seem to be partly projective, and the view that emotional reactions should be subject to criteria of appropriateness not set by our sensibilities alone. Insofar as such criteria of appropriateness would have the same direction of fit as those for beliefs, from world to agent, the degree of autonomy we have with regard to them would be negligible.

On the other hand, since the criteria of appropriateness for emotions may be more relativised to us and our sensibilities than the criteria for appropriateness of beliefs, there might be projection involved in appropriate emotion, too. But as suggested before, the projection involved in emotion may be passive, either with respect to its objects or with respect to those who feel it, or both. Until we have some clearer idea of what the right balance between activity and passivity, projection and "mere" response in the autonomous agent, though, the whole issue remains rather a tangle.

**Autonomy and intersubjectivity**

To this we may add a worry about the relation of autonomy to intersubjectivity. While I have argued for a place for particular subjectivity in moral agency, I have also suggested that one limit on this might be the point where agents no longer share enough common assumptions to make some kind of sense as a moral community. I have also argued that particular subjectivity needs to be subject to some standards of appropriateness. Such standards may in part be set by the constraints imposed by concerns for intersubjectivity in morality.

Intersubjectively permeable emotion may also have more positive roles to play in moral agency, for instance in capacities for empathy and sympathy. And Wollheim-style intersubjective heteronomy might not always be a bad thing. For
that an agent is autonomous does not prove him to be infallible. If his autonomy makes him completely emotionally impermeable to the attitudes of others, it may even preclude necessary readjustments to his moral performance. And in that case, the good of autonomy may be only *ceteris paribus* and not absolute. Or to put it in a more traditional way, autonomy seems to be an executive or instrumental virtue\(^{20}\), one which in a bad agent enhances their vice, as it enhances the virtues of the good agent. This is still compatible with holding that autonomy to be necessary for good moral agency, since those who do right heteronomously may not qualify as good moral agents, however accidentally good their moral performance. But a degree of heteronomy – or at least a degree of interpersonal openness - may also have its uses.

A Kantian might respond to this that the term autonomy would be wrongly applied to a flawed agent, and that autonomy in the true sense is based in reason. But as against this, I have argued, earlier, that we have no sure way of warding off the possibility that reason, as well as emotion, may be flawed. And, of course, some notions of autonomy, such as the existentialist, seem anyway to be favouring will over reason when it comes to constituting autonomy. Also, an agent might be capable of right reason and lack integrity, which suggests reason is not sufficient to be fully autonomous. Or they might be capable of right reason, but be akratic, or just plain fail to act on their intentions.

For any agent falling short of full maturity in their moral agency, then, some degree of intersubjective permeability may not be a bad thing. If someone is genuinely in full possession of all the necessary virtues, of course we would not want them to be swayed from their course by the interference of other and presumably less morally capable agents. But an agent of this kind would be a rare occurrence\(^{21}\). For agents of more limited moral capacities, which is most of us, the

---

\(^{20}\) See for instance Foot (1978); Williams (1985 ch.1).

\(^{21}\) This also points to a connection between autonomy and moral maturity. For an agent lacking such capacities would be like a child: Very possibly capable of spontaneous goodness, but not possessed of capacities to resist sources of moral failure. The less autonomy, then, the more vulnerability to moral luck. Though this might just open a regress about autonomy and moral luck: For capacities for autonomy might not be equally distributed, or even present, in all agents.

The notion of a connection between autonomy and moral maturity also suggests that we might want to speak of possession of both in terms of degrees and more or less complex developmental stories, rather than in either/or terms.
consideration that those who disagree with us may have a point might be usefully allowed.

Nor is fallibility the only reason to allow some degree of intersubjectivity, heteronomy-risks and all, into moral agency. Intersubjectivity may also in itself be thought to be at the heart of morality: think of Kant’s notion of the kingdom of ends, Strawson’s participant reactive attitudes, Hume’s general point of view, and so on. However various, these notions of intersubjectivity do not merely regulate what can count as right moral agency, but are also constitutive of it in more substantial ways. While the intuitions behind this are hard to draw out, they are important, and I will return to the theme throughout the chapters to follow.

**Autonomy and the risk in emotionality**

We should also, I think, resist promoting forms of autonomy which require closing off the capacity for being affected in general. If autonomy requires that we not be capable of being affected, except as we choose to be, autonomy may not do much useful work, and may block out inputs crucial to moral agency. As I argued against the stoic, earlier: what we are affected by and how is a crucial source of information about what is good for us and what is not. And fallibility does not make it less crucial.

That being emotionally open to the world and others brings with it risk, both of suffering and of moral failure, should make for caution, but not for complete emotional withdrawal. I have argued that the aversion to suffering that underlies sour grapes-reactions may have a distorting influence on moral agency, and such reactions were introduced in this discussion as a form of autonomy-failure. But precisely this kind of sour grapes-mentality seems to underlie the stoic’s withdrawal from emotional engagement in the name of increased autonomy.

None of this means that we should allow a free rein to all and any potential sources of heteronomy: on the contrary. For at the least, autonomy serves a useful,

---

22 We could add Aristotle’s emphasis on the social nature of humans, though that raises more complicated questions about whether his praise of the contemplative life as the pinnacle of human flourishing in Book Ten of *The Nicomachean Ethics* is simply a sop to Platonism, given the otherwise very this-worldly and pro-social tone of his argument. For some recent approaches that take intersubjectivity as central to morality, see also Greenspan (1995), Korsgaard (1996).

67
even essential, gate-keeping function in moral agency. This holds both in relation to
other agents and their points of view, in relation to the given of the agent's own
nature, and more generally in capacities for critically examining, and perhaps even
controlling, emotional reactions and how those guide actions. Hence the necessity of
autonomy and other executive virtues. On this point de Sousa (1987:17) notes that
traditional accounts of virtue tend to emphasise, along with dispositions to have the
right kinds of emotions, dispositions enabling us to resist the force of emotional
temptations. De Sousa argues that of the cardinal virtues, prudence, fortitude and
temperance, but not justice, are such emotion-regulating capacities. In view of what
has been said about the need for interpersonal equalisation of participant reactive
attitudes, I am inclined to disagree about justice, though I shall not push the issue
here.

Both autonomy and rationality (which I shall come to next) are also
capacities by which the agent may be capable of managing emotion. I suggested,
earlier, a connection between autonomy and Wollheim's notion of a sense of self. I
have suggested, also, that there may be a connection between the sense of self and
the distinctive features of an agent. The connection will however not be
straightforward. The thin sense of self that Wollheim operates with might be thought
of in part as an abstraction from a thicker sense of self that includes, but is not
limited to, the particular distinctive traits of the agent. The thin sense of self
and autonomy both play a regulating role in relation to that thick sense of self. I shall
return to the thick sense of self in discussing the difference those who feel emotions
make to emotions in Chapters Three and Four.

Ideally, of course, we would want objectivity and autonomy to coincide, and
in practice their respective contributions to moral agency may often be hard to
separate. In many respects, they parallel each other. But it seems that they can come

---

23 Cf. Wollheim (1999: 150): "the term 'sense of self' covers a variety of ways in which a
person can relate to himself. It can cover what the person takes himself to be, or his knowledge of
himself: it can derive from introspection. It can cover his self-esteem, or the particular worth that he
sets upon himself: it can derive from self-assessment. Or it can cover his self-respect, or what he
thinks is due to him as a person, and what he will most probably try to exact from others: it can
derive from his scale of values, and not take his particularity into account."
apart, and that even where they coincide, they do so from different directions of fit and by different standards.

4) Objections to emotion from concerns with rationality

Finally, another commonplace assumption about emotion is that it is irrational or non-rational. These are distinct claims. For if emotion is non-rational, it falls altogether outside our criteria of rationality. If emotion is irrational, by contrast, it falls within at least a broad definition of rationality, but fails to fulfill some further evaluative criteria of success applied to rational phenomena. The question of what counts as providing such criteria is of course a difficult one in its own right, as is the role of rationality in morality. While both are relevant here, the treatment of both will necessarily be superficial.

My view here is that emotion can on the whole be brought within the sphere of rationality. I will also argue that claims for its being irrational are overstated. This applies both to views that emotion is inherently irrational, and to the weaker claim that it is just more often than not irrational. It can still be conceded that there are difficult issues about the rationality of emotion - some of which may be particular to emotion. These points will be examined and developed further in the next chapter, where I discuss current descriptive theories of emotion: Much of the debate about the cognitive nature of emotion turns on appeals to criteria of rationality.

Criteria of rationality

Two kinds of criteria of rationality that are often thought to be required features of moral thinking have already been suggested in the discussion above. One is a requirement on intersubjective consistency. An example of this principle, and the central place it is given in morality, would be Kant’s first formulation of the categorical imperative, the formula of universal law. The underlying intuition is that if agents disagree, that is a prima facie reason to think that someone has failed to exhibit full rationality: competent agents will agree, at least when in full possession of the same facts.
The other criterion is one of consistency across cases: relevantly similar cases should be treated the same. This might be thought covered by the criterion of intersubjective agreement, but arguably this further specification is needed to close loopholes. It is after all possible for a standard to command general agreement without being consistent across cases. Social double standards, for instance, may be shared by a majority of agents, and perhaps even all agents.

Insofar as emotion is often thought to be randomly projective, it might violate at least the criterion of consistency across cases. That it might also be agent-relative in the particular rather than the general sense indicates that emotion may also violate the criterion of intersubjective consistency. I have made some suggestions as to how these charges can be met in the discussion of concerns about emotion and objectivity, above. I will not discuss these particular kinds of rationality-criteria further just here, but they will be relevant again, here and later in the thesis. And it is worth pointing out that these are used as criteria of rationality as well as of objectivity.

**Cognitive and strategic rationality**

Here, the immediate main focus will be two further criteria of rationality that recur in discussions of emotion and also have an important place in assumptions about what constitutes good moral agency. These are criteria of cognitive and strategic rationality. I will briefly set out here how each of these feature in moral agency, and then discuss how each of these criteria applies to emotion. I shall also look briefly at some specific kinds of worries about emotion, such as akasria and ambivalence, which may be thought problematic on more than one kind of rationality-criterion.

The distinction between cognitive and strategic rationality is basically made in terms of the interest that guides their respective criteria of assessment.24 Assessment of a state or action’s strategic rationality focuses on the value of its consequences for our goals and interests. The assessment of a state’s cognitive

---

rationality concerns its relations to the cognitions that enter into its explanation. A state is cognitively rational if these cognitions provide warrant for the state. Criteria of cognitive rationality are usually applied to phenomena like beliefs and judgements, and typically aim to assess success in truth-capturing. Most of the arguments that are brought to bear on the appropriateness or otherwise of emotion tend to draw in some measure on both these, though the notion of appropriateness for emotion is perhaps most often understood as an extension of criteria for cognitive rationality.

**Rationality in moral agency**

So, before we look more closely at the criteria of cognitive and strategic rationality, how do these criteria figure in moral agency?

I take it that strategic rationality will be important in moral agency at least in an instrumental way, and a crucial one, insofar as we aim for our moral agency to be efficacious. However, some want to argue for a more substantive role for strategic rationality in moral agency. For instance, on a utilitarian account, the basic principle for evaluating moral agency is strategic rationality relative to the goal of maximizing utility. On an Aristotelian account, where good moral agency is presented as constitutive as well as causative of the highest good of *eudaimonia*, we may also take it that strategic rationality has a valid and even crucial role in informing moral agency. *Eudaimonia* itself is also defined in terms of rationality, as the perfectioning of the specifically human trait of reason. Even on a Kantian account, strategic rationality may have a substantive role to play: the categorical imperative may be seen as a principle of strategic rationality for the bringing about of the kingdom of ends - of which the imperative is also itself at least partially constitutive.\(^{25}\)

Insofar as we think correct factual perception to be a required quality for efficacious agency generally, *cognitive rationality* will similarly have at least an instrumental role in moral agency. Moreover, criteria of cognitive rationality might

---

\(^{25}\) The distinction I make between cognitive and strategic rationality here may not map simply onto either Aristotle's or Kant's distinctions between pure/theoretical and practical reason. Given the enormity of the task, and the textual issues involved, I will not be taking on this issue here. For more general discussions of the place of practical reason in morality, see e.g. Korsgaard (1986, 1996), O'Neill (1989, 1996).
be thought involved in capturing specifically moral truths. The cognitive rationality required to capture such truths may more or less resemble criteria of cognitive rationality used in other, not specifically moral, contexts. Broadly, the resemblance will vary according to how closely the specific meta-ethical account of value and moral perception in question takes it that non-moral and moral properties and perception resemble each other.

Capturing moral truths, and morally relevant truths, is important for moral agency both as a means and as an end. This suggests cognitive rationality's importance to moral agency is given by ways in which cognitive rationality is strategically rational. Conversely, those elements of agency that strategic rationality typically applies to - actions and motivations - may be shown to be cognitively rational as embodying or capturing some higher moral truth or standard of appropriateness.

Either way, it seems to be a common intuition that morality consists in some form of higher unity of strategic and cognitive rationality. The ideal case for the virtuous agent, for instance, is usually held to be the one in which (cognitive) reason and (strategic) inclination are in accord. Even Kant, who holds virtue of inclination to be supererogatory, thinks it a happier case than that of the merely self-controlled agent. Aristotle, notoriously, denies that the merely self-controlled agent is truly virtuous. Hume’s insistence that the beliefs that accompany the passions must not involve false suppositions about matters of fact might be seen as a criterion of cognitive rationality in morality. That they must also not involve false suppositions about the efficacy of means suggests a criterion of strategic rationality, albeit cast in terms of cognitive rationality.

For emotion, by contrast, strategic and cognitive rationality may come apart: an emotion may be appropriate in cognitive terms and still be less than appropriate in strategic terms. Here, I shall discuss the two kinds of criteria separately first, and then look briefly at how they might interrelate and come apart in the case of emotion, and the significance of this.
The cognitive rationality of emotion

Criteria of cognitive rationality

Starting with cognitive rationality: A state is cognitively rational if we can ascribe to its subject an attitude or belief in view of which the state can be assessed as reasonable. More broadly, a state is cognitively rational if it is arrived at in such a way as to be probably adequate to some actual state of the world that it purports to represent.

The condition for a state to be \textit{minimally or categorially cognitively rational} is that the state is based on an attitude that warrants it. A warranting attitude is one that provides a good reason for the state. For an emotion, the warranting attitude is usually thought to be an ascription to the object of the emotion of a relevant property. Fear, for instance, is minimally warranted by ascribing to the feared object the quality of being dangerous.

A state is \textit{evaluatively cognitively rational} if the warranting attitude is itself warranted, either because it is true (and the subject holds it), or because the subject has good reasons for holding it. A state can only fulfill the evaluative criteria if it fulfills the minimal criteria in at least some vacuous sense.

If a state does not qualify for minimal cognitive rationality, it will be \textit{cognitively arational or non-rational}. If a state qualifies for minimal cognitive rationality but fails the further warranting criterion (2, above) it will be \textit{cognitively irrational}.

The evaluative criteria for cognitive rationality are concerned with success relative to some criterion of appropriateness, such as truth-capturing. Therefore, cases where the subject has good reasons for their cognitive state, but these reasons fail to capture the truth, might be thought to fall short of some more \textit{substantive} evaluative criterion of cognitive rationality for which truth-capturing is required. In this case, the evaluative rationality of the state will be merely \textit{subjective}.

Criteria of cognitive rationality usually also apply across the agent’s attitudes. For instance, if the agent holds some attitude that would defeat the warranting condition of a state, holding that state would be irrational, whether or not
the agent holds the warranting attitude. This also points to the holding of mutually contradictory attitudes as a form of cognitive irrationality.

The agent's responsibility *qua* rational being for any falling short may however be a difficult matter to decide. Assessments may vary with the particular case, as well as with the strictness of any particular evaluative account of cognitive rationality.

In its more substantive form, aimed at truth-capturing, cognitive rationality may also be thought of as a form of objectivity. In its less ambitious forms, though, and notably in its minimal criteria, cognitive rationality is distinct from objectivity.

*Emotion and cognitive rationality*

Claims for emotion to be at least minimally cognitively rational are usually based, as suggested above, on the notion that emotion involves an attitude taken towards a specific object. This attitude, in turn, is to be based on ascription to the object of some specific salient property.

This view has however been attacked. For it seems that we may have emotions without making the relevant kinds of property-ascriptions to the object. In phobias, for instance, the person may claim that they do not actually believe the feared object to be dangerous. But the fear they feel of it may still be intense. In blind emotions, similarly, the person may not think that there is anything about the object that warrants their views. When asked why they hate or love some person, they may be unable to provide any other answers than "I just do"26.

In other cases, the failure to have the relevant kind of warranting attitude may be one of lack of object-specificity: it seems that we can experience emotions, such as free-floating fear, without a clear sense of what they are about. In these cases, emotions shade into moods. And this process may work both ways, often deceptively: some of what we think are our emotions may in fact be moods that have been targeted onto a specific object more or less randomly, and some of what we think are our moods may be emotions that have spilled over onto objects other than the original one.

---

26 For a more detailed discussion of blind emotions and their significance, see e.g. Armon-Jones (1991).
To these examples, we can add cases where emotion fails to occur even in the presence of strong apparent warrant, and cases where emotion is other than seems warranted. While these might more plausibly be seen as instances of irrationality, they could also arguably add up to a cumulative case that emotions are not actually informed by warrant at all.

Insofar as these are seen as anomalous cases, or just in need of further explanation, they need perhaps not worry us. But some authors (Armon-Jones 1991, Nissenbaum 1985) have taken them to point to tendencies inherent to emotion. Armon-Jones (1991) also raises the worry that our tendency to rationalise our emotions, once those emotions are present, makes detecting cases where emotion fails minimal criteria of rationality difficult.

While this point in itself is well taken, however, it seems overstated: cases where minimal rationality appears to break down are often quite detectable, as the examples above show. Also, it seems unwarranted to move from rationalisation covering up failures of minimal rationality in particular cases to a general assumption about emotion. Compare the case of sense-perception: While we may not be able to say for sure at any moment that we are not deceived, that need not compel us into Cartesian-style global doubt. Moreover, such specious rationalisations as do occur are only possible against a background assumption that emotions are usually based in such warranting attitudes. That background assumption can of course itself be questioned, but merely local instances of breakdown in warrant for emotion do not seem to be sufficient grounds to move beyond questioning into assumption.

A possible ground for general doubt about minimal cognitive rationality in emotion is that the criteria for such minimal warrant may be under-specified, and that application of emotion-warranting terms like "dangerous" is open to a significant degree of randomness. This point has already turned up in an earlier version under the heading of projection. While such lack of specificity suggests that we cannot always tell whether warrant is genuinely present or not, however, it does not quite suggest a free for all. I have also suggested, earlier, that some degree of
agent-relativisation and context-dependence provides explanation, if not necessarily justification, for variable application of the relevant terms.

The brute feeling view of emotion-subjectivity also provides grounds for doubt about minimal rationality. For it puts emotion on a par with phenomena like sensations, which are often thought not to have the kind of content that could count as a cognitive warrant. One might however question that assumption even as applied to sensations, which might be argued to be at least quasi-intentional (Rorty 1980b, de Sousa, 1987). I shall argue these points at some greater length in the next chapter: For now, it is enough that we can motivate some initial plausibility for the view that emotion is at least minimally cognitively rational.

*Emotion and cognitive irrationality*

Of course, the argument is not always that emotion is non-rational: the worry may be that emotion is, rather, irrational. Now, if all this means is that some instances of emotion are irrational, that need not worry us. Any kind of state that passes the minimal criteria for cognitive rationality could be expected to have its irrational instances, after all. The concern is rather whether emotion is inherently irrational. Few hold this extreme view - even the Stoics may not have. But there is a tradition of thinking that emotion has inherent *tendencies* towards irrationality, even where it is granted that not all instances of emotion manifest these tendencies.

One example of such problems, mentioned in chapter 1, is that emotion may seem to run ahead of warrant. Emotion may be triggered even where there is only a low and possibly entirely subjective probability that warrant holds, or by imagination (Moran 1994).

Conversely, inertia might be an example of such irrationality. In these cases, the warrant is gone, but the emotion lingers. Where this is due to the agent's failing to capture the change in the situation, emotion can hardly be blamed. But emotion may linger even where supporting beliefs change27. Emotion is often thought particularly prone to such recalcitrance, and cases where the subject, noticing that their emotion lacks warrant, *then* casts about for warrant, are not unknown. Or, failing to make some particular charge against an object of dislike stick, the subject
looks for grounds for other charges. Which brings up again worries about emotion as self-reinforcing and projective. I shall argue, in the next chapter and onwards, that we may in fact help emotion out here by embracing the view that emotions reflect those who feel them: where an emotion reflects an important concern of an agent's, emotional readjustment may only be possible after a degree of change in the agent themselves. And that such adjustment may be a slow matter may have its own uses as a ward against fickleness.

The states of mind involved in *akrasia* are usually thought to involve paradigmatic cognitive irrationality. Of course, *akrasia* may be seen as significantly about as a disjunction between cognitive evaluation and motivation, which may or may not be thought capable of contradicting each other. But we might also understand *akrasia* to involve affectively holding some view that contradicts another view that one holds non-affectively: Aristotle's own metaphor of emotion and reason speaking with different voices comes to mind here. And as argued in the previous chapter, if we abandon the view that reason is always in the right where reason and emotion diverge, *akrasia* may now be seen as an unfortunate side-effect of a useful feature.

Emotional ambivalence, where the subject appears to be feeling mutually contradictory emotions simultaneously, may also be irrational in this way. However, it is not clear that this will work as a blanket charge of irrationality: insofar as an agent is aware of warranting conditions for more than one emotion, ambivalence may be thought to be at least minimally, if not evaluatively, rational (Greenspan, 1988). On the further assumption that one or the other emotion must be more warranted and should win out regardless of any persistent warrant for the other, we have a case against ambivalence, and the further assumption is one that we often make, in this context and generally. However, it is a further assumption, and one that may need further warranting in its own turn.\(^2^8\)

---

27 See e.g. D'Arms and Jacobson (1994, 2001).

28 It is of course possible to get the contradictory emotions apart by some split-level story. The parent who feels both anger and loving relief when a child that has run off and got lost is returned to them may not be directing two contradictory emotions at the same object. Rather, the two emotions may have different objects insofar as the child is seen under a different aspect in each emotion. See Greenspan (1980, 1988), Pugmire (1996). Though even here, we might think that one or the other aspect of the object of emotion should be seen as more salient.
These problem-cases may also raise concerns for strategic rationality, which I will discuss next, before bringing issues of strategic and cognitive rationality together. For now, it should also be said - more on this later - that the degree to which emotion is uniquely problematic in these respects may be overstated. Beliefs, too, may run ahead of warrant, resist adjustment, and present internal contractions. And while sometimes this is due to the influence of emotion, sometimes it is not - and sometimes beliefs are involved in creating problem-cases for emotion. Some similar problems may arise for desires, though I shall have less to say on that issue. Overall though, as I argued when defending emotion against charges of non-rationality, the cases that warrant assumptions of blanket irrationality in emotion seem to proceed by overstating the significance of anomalous cases.

The strategic rationality of emotion

Criteria of strategic rationality

Initially, criteria of strategic rationality might be given as *mutatis mutandis* the same as for cognitive rationality. Thus, the *minimal criterion* would be that a state (e.g. an emotion) or action be held by the agent based on some assumption of the value of that action or state to some goal or interest, where that assumption, if true, would warrant the state or action.

The further *evaluative criteria* would be whether the warranting assumption is itself actually warranted, whether by its truth or by the subject’s having good reason to believe it. The case where the subject’s reasons are truth-capturing would then be one of *substantive strategic rationality*. Failure to fulfill the minimal criterion would make emotion *strategically non-rational*: Failure to fulfill the further evaluative criterion would make emotion *strategically irrational*.

Emotion and strategic rationality

But we should perhaps make a clearer distinction here between strategic rationality as it applies to emotion and as it applies to action. As applied to action,
including action motivated by goals and interests set by emotion, the criteria of strategic rationality seem fairly unproblematic.

The notion of emotion as being minimally strategically rational in being warranted by an assumption that the emotion would serve as a goal to some end may however be rather jarring. For it would seem that this criterion could only apply to cases where the subject deliberately gets themselves into a particular emotional state for a strategic reason. But we usually take it, rightly or wrongly, that such cases are anomalous, and tend to view with suspicion the sincerity of those that come by their emotions in this way. Our intuitions on this point are perhaps not entirely consistent, since we do also at least in some instances seem to advocate such an instrumental approach to emotion: we may for instance say that in a particular situation it would be more useful for someone to feel angry than scared, or to take a positive or assertive stance, both of which are emotion-involving.

This worry about instrumentality remains if we apply the evaluative criteria: Of course, in particular cases anger, assertiveness, and so on, may be warranted strategically, and may be useful to call on, but we would still tend to think that there was something odd, qua agents, about those whose emotional reactions consistently operated in this way. At least this is so if the strategic rationality of the agent is understood to operate, consistently, at the level of conscious choice about particular instances, in the way that we usually assume it does for action. It makes sense to ask, in any particular instance, what one practically and morally speaking ought to do. But for emotion, the “ought” in “What ought I to feel?” may rather be one of applying criteria of warrant - cognitive or strategic - to an emotion already present.

However, the oddness of applying criteria of minimal strategic rationality to particular instances of emotion may belie the sense in applying criteria of strategic rationality to emotion - or particular emotion-types - as such. We may legitimately be concerned with whether our characters are such as to make us likely to react in the right manner, in the right circumstances, towards the right object, for the right reason, to paraphrase Aristotle. Insofar as considerations of the appropriateness of

---

29 Solomon, about whom more in the next chapter, explicitly argues (1978, 1988, 1993) for an account of emotion as action-analogue, on Sartrean-Nietzschean grounds, but in this he is fairly unique among current writers on emotion.
emotions also take into account their relation to our goals and interests, as well as to their being warranted cognitively, we may want to take a strategic approach to emotion in this broad sense. Particularly so if our actual emotions have been liable to undermine our efforts at being good moral agents.

Defenders of emotion have appealed to ways in which emotion might fulfill the evaluative strategic criterion of being *genuinely* likely to promote our goals and interests. Such arguments apply across the emotional range, often in rather surprising ways. Aristotle recommends (albeit finely calibrated) anger as a means to getting and sustaining respect for oneself from oneself and others; Hume and Aristotle both recommend friendship as a constitutive means to happiness and flourishing; Kierkegaard, in *Either-or* (1925) and elsewhere, defends despair as a means to greater clarity about oneself and one’s aims in life.

On the other hand, emotions have also been argued more likely than not to fail evaluative criteria for strategic rationality. Whether a particular emotion-type, emotional disposition or occurrent emotional state serves us well in achieving some purpose will depend on a number of factors: what the emotion motivates us to do, how well it sustains the motivation; its impact on our general capacity for strategic thinking; its effect on our cognitive capacities more generally; its likely efficacy in the particular context of action. Emotion may, in particular instances, affect any or all of these adversely. Or it may help in one way and hinder in another. I have argued that removal of emotion may be crucially demotivating; conversely, though, making good use of emotional motivation may take considerable self-management skill.

These adverse effects may be partly due to the tendency of felt emotion to overwhelm experientially ("swamping"), clouding the capacity of the agent both for skills that seem inherently relevant to moral agency (e.g. calibrating emotional reaction appropriately) and skills that seem more instrumental (e.g. distinguishing

---

For instance, while anger might serve to alert us to some genuine grievance, and motivate us to do something about the grievance, intense anger might cloud or ability to think clearly about the situation, let alone deal well with it strategically. In “venting” anger, for instance, the agent may be focused on their emotional state rather than the grievance that gave cause to it. And the offender may then take umbrage at being the target of venting, and use that as an excuse for not addressing the grievance, either.
what serves in the long term from what serves in the short term). Sometimes, emotion may inhibit the capacity to act at all, as when someone is too paralysed with fear to get out of the way of danger. And insofar as emotion tends to affect both thought and motivation, these too may be contaminated by the flaws in emotion.

In these cases, though, we are dealing with felt, occurrent emotion rather than the emotional dispositions I argued in Chapter One may be doing the brunt of the moral work. Emotion, understood in terms of dispositions across time, and including occurrent episodes of felt emotion, may do its work on the whole fairly unobtrusively in guiding and motivating our behaviours. This is not to say that emotion-dispositions may not also go wrong: clearly they can, and I will be addressing that issue in the chapters that follow. But “irruptive” states (Griffiths, 1997) of high emotional arousal, and the problems they cause, are not the staple of most people’s emotional lives. And over emotional dispositions we have a considerable degree of control: they are also more amenable to deliberate change and to choice than occurrent emotions.31

**Emotion and strategic rationality in agency**

The points about randomness and ambivalence made for cognitive rationality suggested some worries about the overall coherence of the subject qua cognitively rational entity. We might want to introduce some similar notion for strategic rationality. A reasonable meta-level assumption seems to be that there should be some coherence to the agent qua bearer of goals and interests. Otherwise, the state or action that is strategically rational from the point of view of fulfilling one of the agent’s goals may be irrational from the point of view of another goal. *Akrasia*, where the agent acts against what they believe right, also seems to involve strategic irrationality at the level of agency.

The principles by which such coherence at the level of agency would be achieved can be left conveniently vague at this point - some form of internal ranking of goals might for instance be useful, to have a sense of priorities ready in cases where goals conflict in practice. It should however be noted that the degree to which

31 For more detailed discussion of the strategic rationality of emotion, see Frank (1988) and Greenspan (1988).
the agent is capable of exercising such unity of agency may be highly vulnerable to real-world luck. It may also depend on the degree to which an agent’s emotions can be expected to form a coherent structure, a topic I will return to in Chapters Five and Six.

**Some possibly emotion-specific problems about rationality**

Some of the problems discussed above - ambivalence, inertia, *akrasia* - are ones that have been held up as reasons why emotion specifically is problematic for moral agency and agency generally. I have already suggested that the emotion-specificity of these problem cases may be exaggerated: there are plausible equivalents for these problems for other mental kinds such as beliefs and desires. The degree to which these putative problems are necessarily always problematic may also be queried. I have suggested they may all also serve some useful functions. Which is not to deny that these phenomena can be deeply and genuinely problematic.

Insofar as ambivalence leaves us with apparently practically contradictory goals, it is strategically irrational at the meta-level. And this worry about strategic rationality still applies even if we think ambivalence could be cognitively warranted, which suggests that cognitive and strategic rationality may come apart. It may be a moot point whether this is a problem about emotion or a sign of some inconsistency in our normative assumptions about rationality in moral agency.

I have argued, previously, that the possibility of *akrasia* may be strategically useful to us as agents, since reason as well as emotion may go wrong. This again suggests a possible gap between cognitive and strategic rationality as applied to moral agency, though not of quite the same kind that applies to ambivalence. It also suggests that some degree of inconsistency with other mental states and attitudes need not necessarily count against emotion.

Inertia and running ahead of warrant, as well as the occasional opacity or apparent absence of warrant for emotion, may I think all be accommodated as explicable anomalies within an account of emotion that sees it as a *sui generis* form of cognitive awareness. More precisely, I have in mind that the specifically
emotional cognitive work may be done by feeling and feeling-dispositions, which makes the problem cases rather more explicable than they are if we try to assimilate emotions to judgements. The problematic features may also have some degree of strategic value, both for cognitive and strategic levels of agency, which crucially depends on just the same things that make them problematic. These views will be motivated further in the next chapter.

5) Emotion and worries about emotion

To summarise: The kind of charges against emotion considered here raise crucial issues for moral agency: our capacity for correct perception, right reasoning, right motivation, and for autonomy. In each case, though, I have argued that the descriptive element of the charge is overstated, and that the normative element of the charge may itself be in need of some revision.

To objections against emotion from objectivity-concerns, it can be countered that emotion is not just brute feeling, and that its representational content is not purely or randomly projective; nor need emotion be entirely and irreducibly idiosyncratic. And on the normative side, I have argued that we can give at least a qualified defence of agent-relativity, both in the general and the particular sense of the term.

Similarly when it comes to concerns about autonomy: emotion is not altogether beyond our control, and while it can be granted that emotionality makes us vulnerable to externals, the degree of control we have over our emotions suggest this is a manageable worry. However, as already suggested by my argument against the stoic in the previous chapter, the value of such control as autonomy involves may be questioned when it goes beyond a corrective function. So on the normative side, it may not be a bad thing that emotion is not altogether under our control, nor is it necessarily a bad thing that emotion reflects that we are not quite in control of ourselves or our situation.

And finally, I have argued that emotion is at least broadly within the realm of the rational, both cognitively and strategically. As on the previous issues of
objectivity-failure and rationality-failure, emotion is problematic in particular instances rather than on a general level. And again, the fact that emotion can go against reason in its narrowly construed sense may not be a bad thing, a point which calls into question the concepts of rationality by which emotion has been found failing.

Overall then, by qualifying both the descriptive assumptions about emotion involved in the objections and the normative assumptions they make about moral agency, we can make a good cumulative case for emotion being less problematic than its critics have assumed. And as suggested throughout, emotion's critics may also have exaggerated the degree to which these worries are genuinely emotion-specific.
Chapter 3: The difference it makes what emotion is

1) Introduction

Chapter aims and outline

So far, I have not looked very closely into the issue of what emotion is. We are now in a position to examine some current descriptive approaches to emotion, to see what light they might shed on the role of emotion in moral agency, and whether they may help to meet the objections to emotion.

As already stated, consensus in the current literature tends to cluster around analyses of emotion that view it as essentially involving a cognitive element. Within this rather broad notion of emotion essentially involving a cognitive element, there is however a great variety of further specifications. Most cognitive accounts take it that emotion involves appraisal of an (intentional object). I shall refer to these two aspects of cognitive theories as the cognition-claim and the object-claim respectively. A third point which will be important here, and on which cognitive accounts vary significantly, is over how distinctive they hold emotions to be, and how they understand emotion’s differences from and relations to other kinds such as beliefs, desires, sensations and behavioural dispositions.

A common thread in the current literature is that normative concerns about emotion in moral agency may be met by arguing for a cognitive account of emotion. This, I shall argue, is broadly right. However, “judgementalist” cognitive accounts that assimilate emotion to belief or judgement may, insofar as they help make emotion more meta-ethically respectable, do so at the cost of also making emotion qua emotion meta-ethically redundant. And the extent to which such assimilation helps to make emotion more respectable may not turn out to be so great after all. Also, such accounts are descriptively less than fully plausible. I will illustrate this, in Section Two, by examining how a strong judgementalist account may run into trouble, both meta-ethically and descriptively. I will also argue against reduction of
emotion in belief-desire terms, both on descriptive grounds and as regards normative vindication of emotion.

A better approach, I will then argue in Section Three, is to revise the cognition-claim to allow that belief and desire may play significant roles in emotion but to also allow at least a quasi-cognitive role to the distinctively affective element in emotion: feeling. Feeling, as understood here, need not be limited to physical sensations or awareness of physical occurrences, though both of these may be significant elements in at least some emotions. And even where the feeling involved in emotion is physical, I will argue that it is not brute, but rather is at least quasi-intentional and quasi-cognitive. This content of feeling may also be richer than mere binary (pleasant/unpleasant) hedonic tone.

Next, in Section Four, I examine the significance and plausibility of the object-claim. Based on this, I argue in Section Five for a revision of the descriptive account of emotion to emphasise that emotions reflect features of those who feel them, as well as reflecting those things the emotions are about. If we were different, our emotions would be different. So, a number of features that power worries about emotions may have their sources in what those who feel them are like, and what situation those who feel them are in, rather than in emotion as such. Emphasising the degree to which emotions reflects those who feel them may also help towards resolving some worries raised for cognitive accounts by kinds of affect, such as moods and blind emotions, whose relations to intentional objects might be in doubt. Among the human traits that are reflected in the particular form our emotions take are our capacities for creative self-determination and self-interpretation: I argue that these capacities are reflected in the role of what de Sousa calls ideologies of emotion in our emotional learning. These points, and their significance for the place of emotion in moral agency, will be pursued further in the next chapters.
2) Judgementalism and the role of emotion in moral agency

A strong cognitive account - the judgementalist identity theory

I will start by examining how a view that identifies emotion with judgement might fare. This being the simplest, and also the most extreme, version of a cognitive account, how it fares will also shed some useful light on how more moderate versions of cognitive accounts might measure up in relation to the normative issues.

Of the more prominent current authors, probably only Robert Solomon (1973, 1993) and Martha Nussbaum (1993, 1994) espouse this strong view. It is usually referred to as an identity theory - of the relation between emotion and cognitive state, that is. Nussbaum explicitly harks back to the historical Stoics in stating her position. Solomon draws mainly on Sartre, and partly on Nietzsche, for inspiration. Both of them take as their basic position the view that the relation between the emotion and the cognitive state involved in emotion is one of (exhaustive) identity: The emotion just is identical with the cognitive state, and the cognitive state only. By judgement, both authors appear to mean a doxastic state involving truth-commitment to a proposition.

Identity theory and the normative vindication of emotion

Both Nussbaum and Solomon explicitly argue that such a reading of emotion would give grounds for a positive reassessment of emotion's role in moral agency. In this they are both, as they admit, at odds with the normative conclusions drawn by those who inspired their descriptive accounts. The Stoics notoriously took a generally negative view of emotion's role in moral agency. Sartre's (1948) account is less massive in its condemnation, but still casts emotions in a less than flattering light: They emerge from his account as self-serving reactive strategies, rather like

1 I will not be undertaking detailed examination of the history or varieties of cognitive theories of emotion here. For instructive critical discussions of these issues, see e.g. Deigh (1994) and Griffiths (1997). Deigh's article is specifically about cognitive theories of emotion in philosophy. Griffiths' book draws on a multi-disciplinary approach to critique current philosophical approaches to emotion and includes in-depth discussion of cognitive theory approaches in other disciplines, such as the Zajonc-Lazarus debate in American Psychologist, 1984 (Zajonc 1984, Lazarus 1984: see also Zajonc 1980 and Lazarus 1982).
those involved in Nietzschean *resentiment*. Hume, by contrast, seems to take the passions *qua* passions as non-representational, but gives them pride of place in moral agency. So we might suspect, again, that the connection between cognitive accounts of emotion and moral reinstatement of emotion is not that straightforward.

Whether identifying emotion with judgement, or some particular class or type of judgement, will improve its moral status will of course depend crucially on the status of judgements, as well as on the plausibility of the identification itself. In that it runs counter to a number of common descriptive assumptions about emotion, the identity theory is also crucially in need of some error theory about other accounts of emotion with which the identity-assumption conflicts. And there are valid doubts that can be raised about the descriptive plausibility of identity theory on all these points.

Even if the identity theory did succeed on these points, it would not so much give grounds for a positive reassessment of emotion's role in moral agency as make emotion redundant, both as a category and in terms of role in moral agency. If emotion is nothing more than judgement, and its moral-psychological respectability derives just from this identity, then emotion is no longer problematic, but neither is it doing anything that mere judgement could not do.

If emotion is something beyond just judgement, the identity-assumption collapses. And if emotion is some particular *kind* of judgement, then again the question arises of what if any *distinct* positive contribution it makes to moral agency, and how this distinctness relates to these judgements being emotions.

Both Nussbaum and Solomon argue that emotions are *evaluative* judgements. And that would suggest an obvious way in which emotions might be essential to moral agency. The sense of ‘evaluative’, in both authors, is rather broader than the strictly moral. And some of the kinds of evaluations involved, for instance in fear,

---

2 See e.g. Nietzsche (1886, 1887).

3 Cf. Hume (1978: 415): “A passion is an original existence, or if you will, modification of existence, and contains not any representative quality, which renders it a copy of any other existence or modification. When I am angry, I am actually possest with the passion, and in that emotion have no more reference to any other object, than when I am thirsty, or sick, or more than five feet high”. Davidson (1976) argues that Hume has a cognitive theory of pride, and by extension of other passions, but this argument seems to rely on the involvement of accompanying beliefs rather than on what Hume says about the passions themselves. The view that Hume held a cognitive theory of emotions is queried by Baier (1978, 1991).
may not be obviously evaluative in the more usual sense - except insofar as thinking something dangerous is to think it is bad. Of course, we do draw on such broader evaluations in making moral judgements. But it seems that we can make evaluative judgements, moral and otherwise, unemotionally. And we may wonder what light the view that emotions are value-judgements throws on the difference between value-judgements and other judgements. So at least in large part, the challenge of giving a positive account of emotion and its role in moral agency remains the same as it was before the identity theory was introduced.

**The possible benefits of the identity-claim**

Starting with the putative benefits of the identity theory for the normative status of emotion: It may seem that by assimilating emotion to judgement, and more specifically to value-judgement, we can side-step the worries that powered the normative objections. Judgement and the associated faculty of reason: these have not been impugned in the way that emotion has when it comes to playing a role in moral agency. Moreover, in our reasoning capacity we are usually thought capable of objectivity, rationality and autonomy in all the morally important senses of those terms - with the usual qualifying assumptions about our individual or species-relative flaws.

As we have seen, there is a considerable tradition - going back to the ancient Greeks at the least - of holding that it is in exercising our capacity for reasoning and making judgements that we best exemplify our moral capacities. There is also a more recent Kantian-inspired tradition of seeing reason as the core of our moral agency. On this view, autonomy and objectivity, as well as rationality itself, are essentially reason-based capacities. So if emotion is in some sense the same as judgement, then it seems it could have the same moral status. The only mystery would be why anyone should have thought otherwise of emotion in the first place.

However, it is not clear that this approach will work. If emotion is to acquire respectability by identification with judgement, we would need not only to establish that this identity holds, but also that judgement itself actually possesses the respectability that it is intended to lend emotion by identification. What we might call optimistic rationalism, the view that reason's role in moral agency will be a
necessarily positive one, has a long tradition in moral psychology. This often combines with the view that reason, possibly together with some kind of reason-led volition, will also be sufficient for good moral agency, without the need for any kind of affect to be involved.

But, as I have argued earlier, it is not clear that our judgements, or reason itself, for that matter, will necessarily, inherently, possess such respectability. People do after all reason wrongly, and hold false beliefs. Nor is it clear that flaws in reason can all be blamed on emotion. As I have argued, it seems both reason and emotion might go wrong. And for those who wish to use the identity theory to lend emotion respectability by identification with reason, getting reason off the hook by impugning emotion is clearly not a viable solution to cases of failures of reason. Finally, and most crucially, it seems that reason, whether it works right or not, may be insufficient for moral agency: emotion, and affect more generally, may have a crucial and distinctive role to play in informing and motivating moral agency.

Now of course even with these qualifications, reason still has the better general reputation in relation to moral agency as compared to emotion, and even a minor improvement in the status of emotion might be a worthwhile achievement. To claim that emotion is close in nature to mental kinds more amenable to rational adjustment could be a significant gain. Both Nussbaum and Solomon, in identifying emotion with judgement, explicitly intend this to mean that emotions are, at least in principle, amenable to adjustment by argument, and to soundly reasoned arguments at that.

Thus, even if it is conceded that emotion may go wrong, it is also supposed that it may - by the same token, even - be put right. What can go wrong through wrong reasoning may still, at least in those who are amenable to reason, be put right by right reasoning. If having the wrong emotions is just having the wrong beliefs, then by changing the beliefs we can change the emotions.

Moreover, on this account, emotion appears much more under our control. If our beliefs determine our emotions, or our emotions are themselves just (like) beliefs, or both, then it seems we should be able to instill in ourselves the right emotions through argument, as well as change or eliminate wrong emotions already present. So, through reason, we might correct for distorting subjectivity, and (re)gain
autonomy. And if the causes of at least some forms of heteronomy, like the causes of bad subjectivity, lie in faulty reasoning, or at least in faults that reasoning can correct, then most of the worries powering the objections to emotion might be thought met.

Voluntarism about emotion and judgement

A further worry here is however that reason and judgement are also prone to the kind of worries about tensions between objectivity and autonomy suggested when introducing the normative objections to emotion in Chapter Two. This is particularly so in the case of the specific authors used as examples here, since both Nussbaum and Solomon draw on a form of voluntarism to support their view that emotion, as a form of judgement, is under our control.

Nussbaum (1994), based on her reading of the Stoics, offers an account of judgement as assent to a proposition, and argues for our control of belief in terms of the capacity to withhold such assent at will. This is qualified by the admission that this might be a skill that requires learning and practice. One other major obstacle to acquiring this skill is the view that we are just passive with respect to our emotions. In a spirit of optimistic rationalism it might then be suggested that if emotions are like judgements, we should be able to control emotion through the exercise of will - as long as we are also disabused of the notion that emotions are non-voluntary.

Solomon (1973, 1988, 1993), also taking a voluntarist approach to the beliefs he identifies with emotions, offers a more Sartrean account where the decision to believe is made on strategic grounds. (He sometimes offers ‘purposive’ and ‘political’ as near synonyms for ‘strategic’). Casting a situation or object in a particular emotional light is a joint coping- and communication-strategy.

Becoming angry, for instance, may have strategic gains for an agent insofar as this casts them in a flattering light to themselves, the supposed offender, and

---

4 Notably, neither Nussbaum nor Solomon offers any clear accounts of what volition is or how it works.
5 This account of judgment might be similar to the (possibly Stoic-inspired as well as scholastic-inspired) one given by Descartes in The Meditations. Judgement in this case appears as something like truth-commitment to a proposition, where the truth-commitment is made by an act of will that may be suspended - except in the case of necessary truths, which he argues are compelling.
implicit or actual third parties. (‘I am the wronged party here, and I have a right to
desire some form of recognition as such’). It also casts them in a threatening or guilt-
inducing light as far as the supposed offending party is concerned.

On the strategic account, we may also expect there to be a certain amount of
resentment in emotion: Where the originally favoured emotional strategy is
blocked, a reactive second-string strategy will be chosen instead. For instance,
someone whose situation might be thought to warrant anger, but who sees him- or
herself as too weak for their anger to make the desired impact, may instead adopt the
stance of grieved victim - in which a cynical observer might suspect strong elements
of passive-aggressive intent. Solomon suggests that greater clarity about emotion
being a matter of judgement may allow us to avoid such distortions and make more
authentic emotional choices.

Voluntarism about emotion

We might doubt whether such voluntarism about emotion is descriptively
plausible, on a number of grounds. For one, the fact that emotion may be amenable
to adjustment by means of reasoning does not warrant the assumption that emotion is
infinitely adjustable by means of reasoning. Nor need it imply that emotion is a
matter of choice in the first place. Regarding deliberate strategic use of emotion, it
could also be added that this seems to be parasitic on the assumption that emotions
are not a matter of radical choice - though of course a hard-line Sartrean might think
this assumption a matter of collective bad faith. But it does seem rather hard to take
seriously the view that all our emotions could be such disavowed actions. And to
want to make all emotions avowed actions instead seems to be a solution overly
informed by this distorted view of what the problem was in the first place.

Faith in radical choice about emotion also seems constrained by assumptions
that there will tend to be some correspondence between emotional reactions and
salient features of their objects. Also, there seem to be some - albeit broad -
intersubjective constraints on what can count as intelligible instances of an emotion.

---

6 Wollheim, 1999, argues that Sartre’s account of emotion assimilates all emotion to the kind
of reaction-formation involved in “sour grapes” moves. And we might want to resist this kind of
generalization, even where we concede its truth for some emotional reactions. See also, again, Elster,
1983.
Unless emotion permits of something like private language, this suggests that intrapersonal applications of emotion-terms are also constrained by intersubjective considerations. I have argued, in Chapter Two, that both these kinds of constraints may have an important place in warding off worries about bad subjectivity. Which suggests that radical voluntarism may not be entirely good news for the role of emotion in moral agency, either. Unless, that is, we are prepared to jettison either kind of objectivity-concern altogether.

While we have a degree of choice about our emotions, it is usually by indirect means, rather than as a matter of more immediate control or choice about particular emotional reactions. Some of the reasons for this may not be so worrying for the voluntarist. Both Solomon and Nussbaum concede that choices in particular instances of emotional reactions may be restricted by previous choices about what the agent values, insofar as the agent is reasonably intrapersonally coherent. Whether this can be backed up by the further assumption that the agent has radical free choice about his valuations - and his own nature - on some more global level is however doubtful. We might think that people have some reasonable degree of control or choice within the parameters of various social, biological and other givens. We might even think that they could affect, and transcend, those givens to a considerable extent. But that someone could be at least potentially be entirely causa sui in all that pertains to their agency seems neither plausible nor obviously desirable.

**Voluntarism about judgements**

Both Solomon and Nussbaum implicitly base voluntarism about emotion on voluntarism about judgements. But we might also doubt whether voluntarism about beliefs and judgements is entirely plausible, at least in a radical form. We do not normally choose what to believe. Being aware that appearances may mislead, we may examine our beliefs critically. And we are able to revise them. But the

---

7 Insofar as emotions are associated with particular behaviours, they may have strategic uses even where they are not immediately intelligible. Even if people do not understand why you are angry with them, they may not want to cross you. But whether such behaviour is strategically useful, especially when repeated, will depend crucially on social context. And to use emotion in this instrumental way hardly makes a good case for emotion having a place in moral agency.
conditions under which we can do so with conviction are not entirely a matter of
will, if they are a matter of will at all.\(^8\)

We may also have concerns about the normative standing of beliefs chosen
for their strategic value. Normally, we would expect that such moves would be
permissible only under conditions of uncertainty, and only within the area of
uncertainty. Even there, one might have doubts - consider for instance criticisms
directed at Pascal's wager, or at William James' "Will to believe"\(^9\). We usually do
seem to take it that choices about what to believe, insofar as we have such choices,
are constrained by evidence, and should be proportioned to it. To the degree that we
also view *evaluative* judgements as capturing facts that are independent of what we
just happen to think, voluntarism about this class of judgements also seems
troubling. Voluntarism about judgements also seems to open up the possibility of
*akrasis* of judgements. And although this may be a real problem (cf. Rorty, 1983), it
is usually one taken to suggest some failure of rationality. So voluntarism about
judgements, together with the identity-assumption, rather than get emotion out of
trouble, may get both judgement and emotion into it.

**Voluntarism and the objections to emotion**

These points also undermine the claims of the identity-theory to meet
concerns about autonomy, since voluntarism (in whatever degree or form it is held)
seems to be a major part of the identity-theorist's attempt to reclaim objectivity and
autonomy for emotion through rationality. Here, again, there is some tension
between concerns with objectivity and concerns with autonomy. Accept radical
choice and objectivity goes; limit the degree of choice by imposing demands for
objectivity, and autonomy in the form of capacity for choice is constrained. Someone
who embraced a radical individualist form of subjectivism about morality might not
be worried by this tension, since they could abandon claims to moral objectivity and

---

\(^8\) Some concerns might also perhaps be raised, as I suggested in the first chapter, about
whether will necessarily stands in the kind of relation to reason that would warrant combining
voluntarism with optimistic rationalism. Will, as already stated, also seems to be more like affects or
desires than like reason in its direction of fit.

\(^9\) Both cited here from Feinberg (1996), from this volume see also the criticism of voluntarism
made by Clifford (1996); also Williams on deciding to believe in Williams, 1973a.
indulge potentially untrammelled autonomy. But this does not apply to most moral theories.

The voluntarism of Nussbaum and Solomon's identity-theories may not be a necessary feature of judgmenentalism. We can imagine a version that embodies a more passive notion of judgements, and that would not raise these particular kinds of issues. Such a view would on the other hand be left with all the usual worries about passivity, both as applied to valuing, as applied to judgements generally, and as applied to emotions. Insofar as the worrying nature of passivity may be exaggerated, this need perhaps not be a problem\(^\text{10}\). But then, if passivity as such is not a problem, there seems to be less motivation for singling out emotion over judgement on the count of passivity, and less motivation for identifying emotion with judgement in order to lend it respectability.

**The distinctness of emotion - just hot cognition?**

Even if the identity-theory did present a clear basis for raising emotion's normative standing, there is still the worry that there were reasons why emotion in particular was singled out for objection in the first place. And that worry suggests the identity-assumption is itself descriptively questionable. More generally, how does the identity theorist explain that both philosophy and folk psychology tend to see emotion as distinct - even paradigmatically distinct - from belief and judgement? If it is argued that emotions are just beliefs or belief-like states, this may also be construed as a question about redundancy: Why should we need this doubling up? If emotion does have a distinct contribution to make, is there some distinct feature or features of emotion to underpin this distinct contribution? And if so, how does the identity-theory hold up given this distinctness of emotion?

Both Nussbaum and Solomon explain our sense of distinction between emotion and other forms of judgement in terms of emotions being judgements about issues that *concern* us. Emotion emerges from both accounts as something like "hot

\(^{10}\) Gordon (1987, chapter 6) argues that it does not follow from emotions being passive that emotions are states that act on us, or that emotions are involuntary states that are not up to us. Specifically, that we are passive with respect to emotions, in the sense that emotions are products of ways in which we are acted upon, does not prove that the "agent" acting upon us is one with respect to which we are passive.
cognition"; distinct from ordinary, "cold" cognition by virtue of being interested and engaged.

But this contextualization of cognition by interest and engagement is also a key point for negative normative accounts of emotion. The perceived distorting effect of interestedness in particular has been a prime motivation for objections to emotion. Granted, this may partly depend on an assimilation of interestedness to partiality and bias that may not itself be altogether warranted, but that is another matter.

If the hot cognition reading is to work as a basis for a normative defence of emotion, then, it needs to be supplemented by further arguments. We would need to have some account, for instance, of what the positive contribution of hot cognition over just cold would be. As far as some concept of hot cognition actually corresponds to emotion, that point may be answered by the case I made for a necessary place for emotion in moral agency. Emotion, or in this case hot cognition, reflects what concerns us, usually has motivational force, and may be crucial to the kinds of interpersonal attitudes constitutive of moral status. But we would also need some optimistic but plausible account of the chances of avoiding or at least containing and minimising the negative aspects inherent in interestedness. In other words, the challenge is much the same that we started out the thesis with.

As far as the identity theory goes, though, hot cognition may actually be bad news. For now it seems that emotions have features that beliefs and judgements are not usually assumed to have, such as motivational force. They also have at least such distinctness as derives from being contextualised by interestedness and engagement. And even if this distinctness might be thought extrinsic to them *qua* judgements, it strains at the identity-assumption. Thus the hot cognition-account fails to fully answer residual normative problems, and also undermines the descriptive plausibility of the identity-assumption.

Nor will it do for the identity theorist just to appeal to the idea that evaluative judgements, by virtue of being evaluative, must be hot. For it seems that we can hold such judgements unemotionally and disinterestedly - and even in apparent contradiction to our emotions. I have argued in the first chapter that there are good reasons to think that evaluations are globally dependent on our affective sensibilities.
But that need not carry over to the local level so that we should infer that we must be in some hot state in any case where we need to make an evaluation. Again: calm passions may be doing much, even most, of the work.

Also, I have argued for this connection between emotion and moral agency to be based on the specifically affective elements in emotion. And these get rather short shrift from judgementalist approaches. The identity-theorists posit that emotion just is evaluative judgement. But they do not explain why evaluative judgement should be emotion. And there seems to be nothing specifically emotional that is essential to the judgementalist account of emotion. On the other hand, hot cognition looks suspiciously like judgement with added affective tone. So it seems that either the identity-assumption has to go, or emotion is redundant. For if any role emotion can perform could be performed by evaluative judgement, it seems we do not need emotion.

A third possibility is that evaluative judgements, insofar as they are seen as necessarily hot, may by that token themselves be thought suspect. Lyons (1980) for instance, denies that evaluations, or at least the evaluations involved in emotions, are cognitive in the way that beliefs are - in other words, a classical emotivist approach on the lines of Ayer and Stevenson. The evaluations involved in emotions, according to Lyons, have a causal-conceptual basis in perceptions of bodily states of comfort or discomfort. And not only are emotion-evaluations not properly cognitive, but they will also only reflect the agent’s personal standards. Even if we might query Lyons’ particular take on emotion-evaluation, it does at least suggest that the identity-theorist, in appealing to the hotness of emotion understood as evaluative judgement, may not so much have met the normative objections to emotion as run right back up against them.

I will return to this issue of hot and cold cognitions again shortly, below, in discussing further discrepancies between emotions and judgements: for the hot cognition-approach might be thought to provide a way out of cases where emotions and judgements fail to map onto each other as neatly as the judgementalist might hope. As suggested here, however, the hot cognition approach may just compound the judgementalist’s problems.
Some further discrepancies between emotion and judgement

The points made above already suggest strong reasons to think that the identity-theory cannot help us much with the normative objections to emotion. They also suggest the descriptive plausibility of the identity-theory is rather strained. I will outline, below, some further reasons for worrying about the descriptive plausibility of the identity-theory. The first two worries, about direction of fit, and about the way the identity-theory seems to leave out feeling, go directly to the issue of what is distinctive about emotion. But there are also some further points about the nature of the cognitive element that will be worth keeping in mind for revised accounts. Since these will also inform some of what I have to say about desire and feeling, I will go through these points before I outline an alternative account of the cognitive element(s) in emotion.

Direction of fit - a case for desires?

One crucial problem for descriptive plausibility is already implicit in the hot cognition issue. If emotions are typically thought to have motivational content, it seems that they have features that beliefs are usually not thought to have. In this sense, emotion seems to have a desire-like direction of fit, as distinct from a belief-like one - though possibly it also has the belief-like one. This in turn suggests, again, that the cognitive and strategic rationality of emotion may be distinct, with criteria of cognitive rationality applied to its belief-like aspects and criteria of strategic rationality to its desire-like aspects. Which is not to say that these two aspects of the rationality of emotion are necessarily unrelated: the natural hope is that they will typically be interrelated and mutually supporting. So we might think that the judgementalist account needs to be updated by the addition of desires - or at least this will be so in the case of most emotions. It seems, though, that some emotions may not be inherently motivating - a point I will return to in discussing the role of feeling in emotion. And more generally, I will argue that emotions cannot be reduced to desires any more than they can be to beliefs.
Finding a place for feeling

Another element that the judgementalist identity theory underplays is a central experiential feature of emotion: its felt nature. Even where an emotion is standing or dispositional, and not continuously felt, an underlying relation to feeling seems fairly basic to our identification of it as emotional. For instance, friendly feeling, which is perhaps typically a standing or dispositional emotion, will seldom be experienced as a continuous state of occurrently feeling friendly. But if someone's friendliness never manifests as occurrent feeling at all, we would think that they were at the least rather unusual and unfortunate in their psychological makeup. And we might not believe they were genuinely friendly. Further, we might expect that this lack of feeling would make a difference in how such people would perceive, act and be motivated as friends compared to the normal case. Which also suggests, as I shall argue further on, that feeling may not just play an epiphenomenal role in emotion.

Emotion in children and animals

There is also a descriptive problem we could call the 'children and animals'-objection. (Roberts 1996, Robinson 1995). Judgementalist accounts of the Nussbaum/Solomon-variety seem to imply that capacities for conceptual or propositional thought, or both, are necessary for emotion. On this view, what passes for emotion in creatures presumed incapable of conceptual or propositional thought - animals and human infants, for instance - is at best an analogue of emotion, not emotion proper.

This position seems counter-intuitive, though that is in itself hardly a conclusive argument against it. More to the point, it looks suspiciously as if it would only be arrived at as a defensive position against 'children and animals'-type counterexamples. For the similarities between emotional displays in infants, animals and human adults are otherwise fairly striking, as has been noted even well before

11 This is not to say that either Nussbaum or Solomon are necessarily committed to this view. Nussbaum's 1993 Gifford lectures (Nussbaum, 1993) in fact include fairly lengthy discussions of Bowlby's, Klein's and Fairbairn's theories of infant emotions. Some writers, such as Gordon, do seem happy to deny that animals and human infants can have "proper" emotions. And in and of itself, the notion that emotion is judgement does create some problems, as I shall argue below in motivating
Darwin's seminal work on the topic (1896). And that there is some underlying similarity is also supported, admittedly with the usual reservations about the accuracy of self-reporting, by the accounts of human adults when recalling early emotional experiences.

Some writers, such as Gordon (1987), express serious doubts as to whether we can attribute emotions to animals, and argue a reductive account of animal emotion, effectively assimilating it to phenomena like the startle reflex. By contrast, Robinson (1995) argues for similarities and continuities between the startle reflex and the higher emotional processes. The 'basic', pan-cultural emotions argued for by Ekman (1992, 1994 - see also Griffiths, 1997), also appear to have analogues in animals.

**Problems of belief-entailment**

At least in any literal version, the identity-assumption also seems hampered by the challenge of finding specific judgements to be identical with and distinctive of any particular emotion-type (Armon-Jones, 1991). Insofar as this just reflects the fact that rather different kinds of judgements may be involved in people having particular emotional reactions, such as anger, this need not be a great worry to the identity-theorist. They may be able to appeal to some more general judgement that these more varied particular judgements could all be thought to entail. For instance, there are a number of beliefs about what someone has done - lying, stealing, going behind one's back - that might figure in being angry with them. But all these can be traced to some more general description such as "They have wronged me". Though there, again, we might think that the identity of emotion and judgement might come apart - for thinking that someone had wronged one might be thought cause for hurt feelings, sadness or indignation rather than anger specifically. And sometimes - though probably not often - we might not react emotionally at all to such a thought.

Even at this stage, the identity-claim is getting rather stretched. There may however be a more general worry. For it seems that not only could we have the kind of belief that would normally warrant an emotion without having the emotion, but

---

*Revision of how we view the cognitive element in emotion.*
also we may have the emotion without the warranting belief. Someone might feel angry without feeling able to self-ascribe any belief that would appear to warrant the anger. I have argued, earlier, that such cases are anomalous. And dismissing such cases as anomalous would be possible for someone who claimed a weaker link than identity between emotion and judgement. For the identity-theorist, though, the problem is graver.

One way the identity-theorist may try to get out of this problem, indicated above, is to appeal to the difference between hot and cold cognitions. So the identity theorist might try to accommodate the difference between holding a view non-emotionally and holding it emotionally through some appeal to whether the content of the judgement concerned one or not. Nussbaum, for instance, gives an account of felt emotion as caused by the “upheaval of thought” that follows from accepting certain beliefs, such as that a loved one has died. The grief, on this view, is the struggle to come to terms with this need for readjusting one’s outlook in the light of the loss.

However, the process of accepting such an event in thought and the process of accepting it emotionally need not map onto each other in a neat way. On first learning of a loss, some may react with instant emotion; others may go numb. But it is not clear that we can infer from this difference that they also differ in what they believe. Nor is it clear that we can infer a difference in whether the news concerns them. Some broad parallels in the course of thought and the course of feeling might of course be expected. So as we come to a fuller realisation of the implications of the loss, we would expect the grief to deepen, and then gradually recede as we - brutally put - grow used to the loss. But we may still be liable to feel grief over a loss long after it has been accommodated in thought as a fact. (See also Pugmire, 1998: 45-7). And that feeling and thought need not run exactly parallel suggests, again, that the two are distinct phenomena.

A more general point is that emotion, on this picture, looks suspiciously like an irruptive state, a hot upheaval against a general background of cold judgements. This seems to bring us close to the notion of emotions as disturbances to normal human functioning, which is hardly promising for a meta-ethical vindication of emotion. And it is not quite clear how the concerns that make some judgements hot
figure in this picture. The identity theory needs some account of the background of concerns that make some judgements emotional and others not, but such an account seems to be missing. I have suggested, earlier, that Hume’s notion of calm passions might provide an intelligible background for the idea of affect informing evaluative judgements. The passions figure mostly as sensibilities or dispositions, and not irruptively, but a fuller understanding of their nature still serves to explain why we are affected hotly by some things and not by others. Unless such a distinction is built into the picture, however, we may find the hotness of some cognitions but not others rather mystifying.\textsuperscript{12}

These points need not imply that emotions are infinitely underdetermined by judgements - nor do I want to argue such a case. But the degree of underdetermination does seem sufficient to raise real doubts about whether emotions can be straightforwardly assimilated to judgements. This is also supported by there being a degree of overlap between emotions and moods: We do not usually take it that judgements can take, or shade into, such unspecific forms. Though in this last assumption we may be wrong, or at least in need of some further consideration - a point I shall return to in discussing the object-claim, below.

In cases where warranting beliefs appear to be missing, the identity-theorist may invoke subconscious or unconscious beliefs. Solomon, following Sartre, has reservations about the idea of unconscious belief, and claims instead something like disavowed belief, implying bad faith. While this might sometimes be the case, I think we can avoid large-scale imputations of bad faith as long as we do not reject the possibility of subconscious belief.

At this point, a sub-division of the subject \textit{qua} psychological system is often also introduced. So we might appeal to the apparently warrant-less emotion as being

\textsuperscript{12} In Stocker and Hegemann (1996:49), Stocker argues that the Stoics may have had what he calls a "non-dry" account of belief, which is basically a hot cognition account. He bases this on the postgraduate work of one J. Waligore, whose account of the Stoics in his doctoral thesis draws on Nussbaum 1994, Lloyd 1978 and Seneca's \textit{De Ira}. There does seem to be some warrant for thinking the Stoics had an account of rational activity as engaged and interested (see also Sherman 1997, cited in notes to Chapter One here). So the position of the Stoics may be closer to the broadly Humean/Aristotelian approach advocated here than it might first appear. So the upper-case Stoic need not be so much of a problem as the lower-case stoic. But it may still make a difference whether our account of affective engagement with the world is one of hot reason or calm passion. See also Blackburn (2002).
warranted by some belief held by the agent’s subconscious, as already suggested, or to some more or less disavowed aspect of the agent such as their “inner child”.

Attributions of subconscious belief, as well as some degree of subdivision of agents, may both have valid uses in dealing with problem-features of emotion, and I will return to these issues further on.

These need not be problems just for the identity-theorist, either. For more moderate versions of cognitive accounts would typically also want to be able to say that emotions at least entail particular cognitive states, whether we take these to be beliefs or not. We would presumably need such a notion of entailment, for instance, in order to have some idea as to what could count as a warrant for an emotion. This would particularly be a concern if there is an apparent lack of warrant, and we want to see whether we can, for instance on the basis of behavioural clues, attribute some subconscious belief that could serve as a warrant. For if we cannot attribute some warranting state to the person who has the emotion, it seems we may have to give up on minimal cognitive rationality. Loss of minimal rationality is a particular worry insofar as emotions, even ones with no apparent warrant, influence subsequent beliefs, desires, actions - and moods and emotions. I think these problems can on the whole be solved - but only by readjusting both the cognition-claim and the objectual claim considerably.

3) Some ways we might want to revise the cognition-claim

In view of these further problems for the strong judgementalist approach, it seems the cognition-claim is in need of some degree of revision from its judgementalist identity-theory version. The suggestions I will make here are broadly compatible with the view that some instances of emotion will be specifiable in the kind of terms preferred by the judgementalist. My claim about judgementalism is

---

Armon-Jones (1991) suggests differentiation with respect to an object in thoughts and behaviours as the main way in which we identify the object of a subconscious emotion. We can tell that someone subconsciously hates, loves, fears etc. some other person by the way that they think and act (including speech acts) with respect to that other person, even where the emotion is otherwise disavowed or covert. The nature of the thoughts and behaviours indicates the nature of the emotion, as
only that it will not work as an overall approach. Based partly on these suggestions, I will also go on to make a case for a reading of feeling as at least quasi-cognitive. Finally in this section, I will look at some issues about the place of desire in emotion, and how this might tie in with feeling.

**Strong versus weak senses of cognitive**

As a first qualification, then, the notion of *cognitive* that the judgmentalistic identity-theorist operates with is too strong. One way towards this is the ‘children and animals’ problem. But even for adult humans, emotion need not entail some doxastic state such as judgement. The thought-content involved in emotion may not be more than a construal, or seeing-as. (Armon-Jones 1991, Roberts 1996). Following Armon-Jones, I will call this latter notion of cognitivity *weak cognitivity*.

If we include among emotional reactions the startle-response or phenomena like it, we might think that the thought-content of emotion could at least in some cases be virtually nil. Or at least, the thought-content of emotion need not always be thought of in terms of higher levels of cognition that can be cashed out in conceptual or propositional terms\(^\text{14}\). Maybe not even in verbal terms\(^\text{15}\). Often, emotions seem to elude definition in precise, literal language. More figurative uses of language may however do a good job of capturing emotions even of this kind\(^\text{16}\).

---

\(^{14}\) The differentiation indicates the object.

\(^{15}\) I will not be dealing here with the specific issues raised by attempts to cast emotions as propositional attitudes (see e.g. Thalberg, 1977). While I take it that many, if not most, instances of felt emotion in human adults can be rendered in terms of “feeling \(F\) that \(p\)” where \(p\) is a propositional object, emotions may have only vaguely defined intentional objects, let alone propositional objects. That emotions, as I shall argue below, may have only weakly cognitive grounds also suggests that formalising emotion in this way may be a forced translation exercise rather than a precise way of capturing the actual cognitive content of the emotions of real agents. And that is before we even begin to consider how murky the cognitive content of child and animal emotion may be.


None of this is to deny that the emotions of human adults, at least, may involve "high-level" cognition, and high levels of conceptual complexity and sophistication. And this may be an important reason why human adults are more capable of managing their emotions by cognitive means than children and animals. Though by the same token, they may also be better at specious rationalisation, which may also have distorting effects on the capacity for genuine emotional reactions. On genuineness of emotion, see e.g. de Sousa (1987, on the "paradox of the phony") and Pugmire (1994). This issue will be focused again in Chapter Five.
Cognitive scheme rather than state?

We might also think, as Rorty (1980b) and de Sousa (1979, 1987) suggest, that the cognitive element in emotion need not be a specific cognitive state, but may be something more like a loose conceptual scheme. Specifically, Rorty argues that the cognitive element in emotion may involve "prepropositional but intentional habits of salience, organisation, and interpretation" (1980b: 104). Emotions of particular types might then be thought entailed when their relevant pattern of salience applies. The criteria for this might, or might not, be specifiable as specific beliefs. At the least, the criteria would be highly context-sensitive.

For the application of an intentional scheme that may not be easily formulated in terms of beliefs, Rorty offers the example of someone surveying a landscape from the point of view of military defensibility. Here, we may not be able to capture the cognitive richness of their perceptions fully in terms of individual specific beliefs. Though clearly some such beliefs would be expected to follow from the application of the cognitive scheme - for instance beliefs about where it would make most sense to position one's forces. Emotion may involve something analogous.

This allows for appropriate occasions of a particular emotion-type to be rather various, since different parts of the pattern of salience may be applicable, and in different ways, depending on the particular case before the agent. It also allows for there to be some degree of overlap between schemes, since some of the same features may figure in more than one scheme. That is not to say that the features will figure in the same way in one scheme as they do in another: I might view a particular feature of someone's behaviour towards me somewhat differently when considering it as cause for sadness than when considering it as cause for anger. In at least some cases, I might reasonably think both emotions warranted - which suggests one way that ambivalence and mixed emotion could be at least minimally rational.

In view of the animals and children-objection, we might also want to specify that the cognitive schemes involved in emotions may just be proto-conceptual or quasi-conceptual.
Quasi-intentionality

In "Explaining emotions", Rorty (1980b) also introduces the notion of "quasi-intentional sets that can, in principle, be fully specified in physical or extensional descriptions". Such quasi-intentional sets "form patterns of focusing and salience without determining the description of that pattern". While the quasi-intentional set can usually be given an extensional definition, in other contexts "particularly those that move from functional explanations toward interpretative or rational accounts, the intentional description is essential" (1980b: 114). A broadly similar notion of quasi-intentionality is also used by de Sousa (1987, ch.4) in discussing those elements in emotion that come closest to being like animal instincts.

In Rorty's example, for instance, the character Jonah has developed a quasi-intentional set that is responsive to perceptions of what he construes as overbearing behaviour on the part of female authority-figures. This quasi-intentional set was developed by Jonah in response to early imprinting from two sources: His mother's behaviour, which he experienced as aggressive towards him, and disapproval of his mother indirectly expressed by his grandfather, who served as one of Jonah's main role models. As a result of these unfortunate formative influences, Jonah has developed a 'magnetising' interpretative disposition, which manifests itself in an irrational resentment of his boss, Esther, and distorted beliefs about her.

The quasi-intentional set that underlies Jonah's resentment of Esther is responsive, or rather makes Jonah responsive, to cues from facial displays, tone of voice, stance, in his perceptions of Esther. Given the distorting effect of the magnetising disposition, unfortunately, Jonah tends to misinterpret what he is picking up. Cues that for others would suggest only a woman in authority, and one who listens to others at that, suggest aggressiveness and an overbearing attitude to Jonah.

Now, quasi-intentional sets need not have this distorting effect. And the cues to which such sets are responsive may play a fundamental role in our affective perception of the world. This is particularly true of interpersonal cases, since such cues are crucial in our assessments of the emotional states of others, which in turn may trigger reactions of our own. Smells, tastes, sounds and visual and tactile
impressions also trigger emotional reactions and memories in relation to non-personal parts of our worlds. Often, perhaps even usually, such processing takes place at a non-conscious level. Sometimes, as with Proust’s *madeleine*, the subject may be fully aware of how the sensory cue triggers the emotion, and why. And in many cases, if not in Jonah’s case, information gathered from such sources is crucially useful to us.

**Cognition-claims and the place of the body in emotion**

It also seems that in at least some emotions, the subject has "quasi-representational" states at a physical level. Changes in the subject’s body state involved in emotion, such as pleasant or unpleasant sensations, may have at least some *quasi-representational* content, and not necessarily in a merely “boo-hooray” sense. “Butterflies in the stomach” for instance, or a “sinking” feeling, are not just assessable for pain/pleasure valence, but have qualitative distinctness beyond that.

Goldie (2000a: Chapter Three) suggests that when the body is involved in emotional experience, perception and expression, we may talk of it as "borrowing intentionality". This may help take away some of the worry that has been raised about more overtly physicalist accounts of emotions in terms of sensations or awareness of physical change: notably, William James’ much-contested and possibly misunderstood account in "What is an emotion?" (1884). If we allow the body to borrow intentionality, we may view the physical processes involved in emotions as instances of the way the body is part of the larger intentional system that is the agent or person.

For those emotions whose correlations with physical sensations and awareness of physical change are fairly straightforward, such physical underpinnings are often crucial for self-ascription of the emotion. I may become aware of being afraid, for instance, through the physical manifestations of that fear, even where I have otherwise blocked my fear from awareness. This also allows for some degree of emotional self-management through biofeedback. Deliberately breathing calmly and relaxing tensed muscles may for instance be an efficient way of reducing feelings of fear - as long, that is, as it can be sustained against renewed waves of fear-signals.
from the body. Similar mechanisms also tend to underlie the functions of psychotropic medication.

Something like this method of emotion-management may also be used for higher cognitive emotions, such as gratitude, optimism, and interest. While these do not have such specific physical correlates, they do have behavioural ones. Consequently, these emotions may also to a great extent be induced through use of acting as-if. And of course we may also often, if not always, be able to induce emotions, change emotions, and even bring emotions to a halt, through thoughts and arguments. But sometimes, emotions are resistant to arguments - possibly because the mental level at which their subjective warrant is held is not easily accessible to arguments, as in the case of Jonah's distorted quasi-intentional set.

If we allow for borrowed intentionality, we may be able to bring emotion-management at more brute biological and behavioural levels closer to cases where emotion is managed by more overtly cognitive means. This might also take away some unease about the notion that these methods of adjusting emotions are necessarily brute and non-intentional.

The borrowed intentionality of the physical also makes cases where mere physiological change is mistaken for emotion intelligible. For instance, in Schachter and Singer's (1962) experiments, subjects were given adrenaline injections, and then exposed to stimuli likely to elicit aggressive responses - specifically, obnoxious behaviour by a stooge planted by the experimenters. Those subjects that were informed about the adrenaline content of the injections tended not to self-ascribe anger, or even irritability, despite the physiological arousal experienced. Those who were not informed tended to self-ascribe anger elicited by the provided "target", the stooge.

Such cases should not, I think, be taken to show that we are misled by the body into thinking non-intentional, non-cognitive states are intentional and cognitive. Rather, we may assume a broad background connection between characteristic patterns of physical arousal and the borrowed intentionality of the body. Mistaken self-ascription of emotion is due to local breakdown in this connection. And the misleading character of such local breakdowns is parasitic on the overall background connection. Awareness of this possibility should make for
caution, and may often make for grounds to examine the aetiology of our emotions carefully. But to take it that such local breakdowns show a lack of overall connection with intentionality for such states seems unwarranted.

The place of feeling in emotion

With these adjustments of the cognition-claim in mind, I now want to suggest that the most emotion-specific element in emotion, namely feeling, is at least quasi-intentional, and possibly cognitive in some stronger sense. By feeling I mean not just physical sensation or awareness of physical occurrences, but also what Stocker (1983) calls the "psychic feeling" of emotion, the distinctive qualitative experience of emotion.

Qualitative and quantitative variations in feeling are crucial experiential and distinguishing features of specific emotional episodes. But the role of feeling goes beyond helping us identify what emotion we are having. What we feel, and how strongly, also has an impact on how we think, desire and act - at least for anyone with less self-control than Mr Spock. And how this impact manifests itself depends crucially on what particular feelings we have - fear and anger, for instance, express themselves in clearly distinct ways. So feeling is not just epiphenomenal.17

This might perhaps raise further worries, if feeling is understood as brute, non-cognitive or non-intentional. However, I would argue, neither physical nor psychic feeling is really that brute. Both have at least quasi-representational content. And both are capable of subtle qualitative as well as quantitative nuances. This holds even where we consider only the hedonic valence of emotion, its pain/pleasure-content. Anger, grief and fear are not painful in the same way. Nor are hope, happiness and gratitude pleasant in the same ways.

Often, there is also some mix of painful and pleasant feeling in emotion. There may be an element of pain in hope, from the possibility of disappointment. The pleasure of happiness may be mixed with some pain at the fragility of happiness - and that pain at fragility may even serve to emphasise the pleasure of the actual happiness. Grief may have elements of pleasure when we think of the lost one, as

---

17 See also Leighton (1984), Stocker (1987), Ridley (1997), for discussions of how feelings
well as pain at their no longer being with us. Some emotions, like nostalgia, may seem to be mixed by definition. And as these examples suggest, the “mix” may not be accidental: Both painful and pleasant feelings, then, bear on what is salient to the emotion, even where they do not bear on exactly the same aspects of it. And this relation to salience indicates at least some quasi-intentionality to feeling.

Some feelings, such as that of surprise, do not seem to have any definite inherent hedonic tone, and some cases of emotion, though they are felt, may not have hedonic tone at all. Feeling surprised might be pleasant or unpleasant - or neither. Pugmire (1998: 65-71) uses this point to argue against what he calls the “hedonic-binary” account of feeling: the view that affect always has hedonic tone, which in turn is either pleasant or unpleasant. Feelings directed at the unfamiliar may be “keen and yet, at least in the moment, hedonically neutral”. Emotions to which this might apply include, apart from surprise, awe, puzzlement, feeling weird, feeling overwhelmed. Some first-time experiences may also be hedonically indeterminate in this way - we may not be sure if we like them or not.

While, as Pugmire concedes, this does not change the impression that the overwhelming majority of feelings have hedonic tone, it does indicate that hedonic tone, binary or not, is not all there is to affective tone. And this point can be emphasised without losing sight of the fact that the hedonic tone of emotion may be crucially important for the way emotions motivate and inform us as agents.

Feelings, then, are capable of fairly complex and subtly modulated content, which is at least quasi-cognitive. And if we allow this, we may, crucially, be able to explain the cognitive nature of emotion at least partly in terms of the cognitivity of emotion’s specifically affective element. And this suggests a way in which the cognition-claim may be used for meta-ethical vindication of emotion without having to give up on the idea that emotion is a distinct form of cognitive awareness. The worry about giving up on the idea of emotion as distinctive arises in the first instance from the descriptive consideration that it seems to have significant features that other
cognitive kinds do not. But as far as normative concerns go, it also seems to be a worry to give up on distinctiveness: As we have seen, rehabilitation without distinctiveness, if it works at all, would leave emotion *qua* emotion redundant.

That feeling is cognitive at least in some weak sense is also supported by the fact that affective tone is usually highly responsive to adjustments of its underpinning by more conventional kinds of cognitive states, that is, by those cognitive elements in the emotion that are not specific to emotion. This connection is often expressed in terms of affective tone accompanying or being directed at cognitive states, and this seems to me broadly right. But as I have argued above, the accompaniment is not so consistently exact as to suggest identity. And the thought-feeling connection may go in the other direction: Feeling may direct thought. And, not least, feeling may, as I have suggested here, have some cognitive complexity of its own.

I do not want to argue that emotion can simply be identified with feeling. For one, any particular emotion may involve a complex mixture of different feelings. And some of the same feelings may feature in different emotions. Combined with more or less motivated selective attention, this sometimes leads people to misdiagnose their emotions. Nor do I want to say that beliefs, or their weaker cognitive analogues, can be left out of the account of emotion. If we did not have some minimal belief-like awareness of the world, however inchoate, there would be nothing to have emotions about in the first place. Desire, which I shall come to next, also plays an important part in most emotions. But I do want to argue that there is a specifically emotional element in the cognitivity of emotion. Insofar as this specifically emotional element is typically directed at other cognitive content, such as that of belief, that is already present in the mind, we might think of emotion as a

hedonic-binary model, see also de Sousa, 1999b.

Sherman (1997:54) cites as an example a poem of Sappho's: "When I see you, my voice fails/my tongue is paralyzed/a fiery fever runs through my whole body/my eyes are swimming/and can see nothing/my ears are filled with a throbbing din/I am shivering all over..." and notes that "Literary history, social convention, and perhaps evolution conspire to tell us this is love. But even here it is not hard to imagine that what is described could be dread or awe, or perhaps mystical inspiration." And while this example goes specifically to physical feeling, similar points can be made for "psychic feelings". De Sousa (1987) citing the same passage, suggests that out of context, the emotion described might even plausibly be anger or hatred.

Elster, 1999a, for instance, argues that Jane Austen's *Emma* manages to mistake feelings of
meta-mode of awareness. It might also figure as a meta-mode in that it directs and shapes the way we pay attention to other kinds of cognitive content.

**Emotion as background disposition versus as occurrent feeling**

A final point about feeling here: In arguing for a role for feeling in emotion-cognition, I do not just mean occurrent, consciously experienced feeling. I take it that feeling, as well as other mental kinds, may occur in non-conscious, subconscious or pre-conscious forms. This might be supported by the fact that we can, demonstrably, fail to notice what we are feeling at the time we are feeling it, and only later realise that we were afraid, or angry, or in love. And like belief, I take it that feeling need not be conscious to play a role in directing attention and shaping interpretations. The feelings that shape our emotion-cognitions may be to some extent inchoate.

Most of the time, in fact, affective dispositions, such as standing emotions, may be doing the work, rather than occurrent feeling\(^{21}\). Loving someone, for instance, is rarely felt as a *constant* conscious feeling. But even as a merely standing emotion, love still shapes our beliefs, desires and actions towards the person in ways that they would not be shaped if we did not love the person. Given that felt emotion may occupy the mind to the point of excluding from awareness anything that seems irrelevant to the particular emotional state, we might also think it a good thing most emotion takes place in the background, psychologically speaking.

**The place of desire in emotion**

Insofar as a main obstacle to reducing emotion to belief is a problem about direction of fit, there may be a temptation to try to reduce emotion in terms of desire instead. This might be supported by the notion of emotion as projective. And that in turn points to assimilation to desire not being a very good thing for emotion’s meta-ethical standing: much the same worries have been applied to desire as to emotion. I am inclined to think that the worries about desire, as well as about emotion, are

---

boredom for signs of being in love.

\(^{21}\) Similar lines of argument are made by Michael Stocker (e.g. in Stocker and Hegemann 1996), Richard Wollheim (1999) and Peter Goldie (2000b).
overstated. On the whole, both are usually at least minimally intelligible, and we have some idea of the criteria by which they can be shown to be so. We are not surrounded by people who "just want" saucers of mud and have no idea as to what description these are desirable under (cf. Anscombe, 1957).

Defending desire is not the issue here, though. But a few points about the place of desire in emotion seem in order here. For there is at least some broad, general connection between emotions and desires. Most emotions entail characteristic desires: this, as well as the hedonic tone of feeling, accounts for the motivational role of emotion.

But emotions are not reducible to the desires they involve. There is more to fear, for instance, than the desire to run away: Overcoming the desire is not yet to overcome the feeling, even if it is an important step in the right direction. And what desires are associated with an emotion may vary, too: some but not all conceptions of anger insist that a desire for revenge follows from it. Desires of particular kinds may also feature in more than one kind of emotion, and we would still be able to distinguish the emotions. The desire to lash out at someone, for instance, may figure in anger, in hate, or in malice. And although those emotions have some important similarities, they are still distinct - and need not entail each other. Moreover, emotions inform desires, so that the otherwise same desire might express itself in significantly different ways depending on the emotion involved.

Insofar as most emotions have hedonic tone, and may by that token be viewed as pro- or con-attitudes (see also Hamlyn, 1978), emotions bear a family resemblance to desires. And as I have argued earlier, it may be important that desires are affective. But that does not mean that emotions can be reduced to desires. For one thing, as argued above, it seems that there are emotions that lack specifiable hedonic tone, but are nevertheless felt.

It may be that we can explain the lack of hedonic valence in the kind of emotions adduced by Pugmire, above, as a result of their being directed at the unfamiliar. We may just have stumbled outside the area that we have a hedonic map for, into areas where we do not know our preferences - yet. And in this regard, these emotions may be anomalous. But that still indicates that emotion, as a mode of interested awareness, may extend beyond pro- and con-attitudes even if those make
up most of the terrain. And even emotions that do have definite hedonic tone do not necessarily entail desires. Being happy, for instance, one might not want anything in particular - though one might have an implicit or meta-level desire to continue in the happy state.

Desire may also figure in emotion as an antecedent, emotion being the reaction to the perceived or expected satisfaction or frustration of desire (Gordon 1987, Wollheim 1999). This again seems to run into the point about some emotions being non-hedonic. Moreover, for those emotions that have hedonic tone - the majority - ascription of antecedent desire may still be a moot point. When we react positively or negatively to something, we might think that all that is going on is that we are discovering what sort of things please or pain us, not that we are uncovering some inchoate antecedent desire that has been satisfied or frustrated. This however leads into deeper issues about the nature of desires and their distinction from 'mere' hedonic tone than fall within the scope of this thesis.

Insofar as we might worry that emotions arise out of and give rise to bad desires, that would also give cause for worry about emotion. But insofar as we can give an account of emotion as a broadly rational phenomenon, worries about the desires that it gives rise to may be put at rest. And insofar as antecedent desires give cause for worry, the source of the problem seems to be the desires rather than emotion as such. Both of these points also suggest that the larger context of worry, both about emotion and about desire, may be about human nature and the motivations inherent in it. And that is a rather larger issue than I mean to take on in full here, though I shall have something more to say about it through the next chapters.

---

22 This is not to deny that some emotions might be seen as fairly straightforwardly following from antecedent desires. For instance, I may be angry because the neighbours are noisy, frustrating my desire for quiet, and be happy when the noise stops. But this need not suggest that I was angry in the first instance because I started out with a desire for quiet. I might have developed that desire, at least in its more than inchoate form, out of the discovery that I really did not enjoy being disturbed by the neighbours.
Weak cognitions and minimal rationality

Before going on to the specific problems raised by worries about whether the object-claim holds up, then, some stocktaking might be in order. In view of the revisions to the cognition-claim, what kind of case can we make for the rationality of emotion? I have argued that the kinds of cognition involved in emotion, including the specifically emotional elements of such cognition, may be only weakly cognitive, possibly taking the form of something like a loose conceptual scheme or a quasi-intentional set. I have also argued that the physical states involved in emotions may have borrowed intentionality, and may have some quasi-representative content.

This should however be compatible with claiming minimal rationality for emotion. We might claim physical reactions, weak construals, and seemings, as warrants for emotions. If asked why we are afraid of someone, it makes at least some minimal sense to appeal to their presence making us feel physically or psychologically uneasy, or to some more or less easily demonstrable pattern of threatening features on their part. Such appeals might even have some claim to conferring evaluative rationality. For the perceptions that we have through weak cognitions, conceptual schemes and quasi-intentional sets may be appropriate enough.

That the various weak forms of cognition involved in emotion may also be rather more inchoate than judgements also goes some way towards explaining why we have cases of apparently absent warrant. For we may not always be conscious of the way in which we pick up on salient features: So the warrant might be there, just not fully present to consciousness.

And it may even make considerable sense not to be conscious of most of one’s cognitive states and processes at any particular point. As long as they are reasonably functional, we may as well let them get on with making their contributions without insisting on full disclosure at every step of the way. We might appeal, at this point, to the usefulness of something like a computational unconscious, whose states and processes are not necessarily accessible to verbal report and introspection. Though it seems worth noting that the computational unconscious might come with its own costs. Griffiths (1997:153-4) cites Rey’s
(1988) use of unconscious processing to explain *akrasia* and self-deception, which again suggests these problems may just be bad side-effects of otherwise useful mental arrangements.

What might give grounds for worry is if the states and workings of the computational unconscious are *inherently* inaccessible to introspection. And even if such radical impenetrability is not a real worry, we may have concerns about the demonstrable difficulties of gaining greater insight into the hows and whys of emotions that have gone wrong. In the case of Rorty's Jonah, for instance, we might have to rely on some fairly indirect methods of inference to track the sources of his resentment of Esther. But in principle at least, some such tracking usually seems possible.

Readjusting the emotional reactions into some more appropriate form would of course be a further challenge, but again there seems to be some grounds for hope that, through argument or more indirect means, we might succeed.

A final point to be made here relates to an argument I made earlier: it is not always clear that our higher cognitive levels are the ones that get it right. So there may be some value, albeit involving risk, to emotion's being recalcitrant to apparent defeating conditions. Someone might at higher levels of mental processing rationalise themselves into a false belief, and the presence of a countervailing emotion may serve to undermine that belief.

So Himmler, as Bennett (1974) argues, suffered from a sense of being morally flawed because he felt a compassion for Jews that seemed irrational in the light of his belief that their eradication was logically and morally necessary for the higher good. And we could imagine how Himmler might try to argue or habituate himself out of feeling this compassion, possibly with some success. So any optimism generated by the consideration that emotion can be brought into line with reason needs to be tempered by this opposing consideration. Some emotions that appear irrational to those who have them, and very possibly also to their environment, may in fact be right, and reason in the wrong. We might think that Himmler had just got reason and emotion the wrong way around here - that his attachment to Third Reich ethics was an affect-based mistake, and his compassion was in fact the voice of right reason. But the possibility of such mistakes, by itself, would give us some cause for
doubt about whether we are always automatically right to try and bring emotionsound to what we see as right reason.

4) The significance of the object-claim

Turning now to the other main aspect of cognitive theories: The object-claim. The kinds of warrant that cognitive accounts think can be invoked for emotions rely crucially on the role object-directedness and object-appraisal are assumed to play in emotion.23 The object-claim combines with the cognition-claim to give us an account of emotion as a mode of intentional awareness. Under at least a Davidsonian assumption about rationality, intentionality is also a requirement of considerations of teleological rationality being brought to bear.24 The telos in question might for instance be truth-capturing, or the promotion of some goal of fitting the world better to my tastes and interests. Hence, respectively, cognitive and strategic issues of rationality.

So far, I have examined only the cognition-claim in any depth. Here, I will set out first how the object-claim and the cognition-claim together make up a case for the cognitive and strategic rationality of emotion. I will then look at some worries that have been raised about the object-claim. I will argue that these need not provide a strong case against the basic rationality of emotion. However, in defending the basic rationality of emotion it may be necessary to appeal to features of the subjects of emotion as well as to those of the objects.

---

23 This somewhat simplified account broadly draws on that given in de Sousa (1987, chapter 5), and in Armon-Jones (1991, ch. 2). There are a number of further complicating features of the attempt to give an account of emotion as a form of cognitive object-appraisal that I will not be dealing with in depth here. Nissenbaum (1985) aims to deconstruct the notion of emotion as object-appraisal as incoherent and multiply ambiguous, rejecting the accounts of for instance Kenny (1963) and Wilson (1972). Her claim seems overly based, however, on considerations about reference-failure and opaque intensional contexts that also apply to for instance beliefs. Armon-Jones argues from moods and blind emotions that the objectual claim may be in need of radical revision. Insofar as both authors' arguments seem overstated, I will not be engaging with them in full here. I think worries about moods and blind emotions can largely be met on the lines I shall outline briefly below.

24 The warrant for this requirement of intentionality comes from broader considerations about explanation of mental events and their directions of actions discussed in Davidson, “Mental events”, “Actions, reasons and causes” and “Causal relations” (all in Davidson, 1980a). I will not be examining these particular issues in depth here.
Object-appraisal and the cognitive rationality of emotion

Starting with the objects: following Hume, a distinction is usually made between the object of an emotion and its cause. The object of the emotion is the intentional object of the emotion, the thing, person or state of affairs that (barring reference failure) the emotion is directed at. Hence, the object of an emotion is sometimes referred to as its target (de Sousa 1987, Armon-Jones 1991). The cause of the emotion is the feature or features attributed to the object which make it the object of that emotion: Being dangerous, for instance, in the case of fear.

For cognitive rationality, the connection between the cognition-claim and the objectual claim is this. The features of an object that provide the cause of the emotion are also those by which the object qualifies (or not) as an instance of the formal object of the emotion.

The formal object of an emotion is the general description under which an intentional object qualifies as an object of a particular emotion. The formal object thus also defines the particular emotion. The formal object of fear, for instance, is, loosely defined, the dangerous or threatening. The formal object typically features in the emotion as an implicit second-order characterisation of the target-object and its salient properties. Broadly, the formal object defines putative targets of the emotion-type in question as viable or non-viable, compelling or non-compelling (Armon-Jones, 1991). The object's being an aggressive and heavily muscled person might

---

25 The distinction between objects and causes of emotions is however prefigured by for instance Aristotle, who distinguishes between what people feel emotions towards, and on what grounds (The Art of Rhetoric, book 2, 1378a). Aristotle however also includes “what those are like who have the emotions” as a crucial element in the analysis, a point I shall draw on here in the later discussion. Kenny (1963) also refers to the use of notions of objects and causes of emotions in the Scholastics, and through Scholastic influence in Descartes.

26 This may also have some explanatory bearing on the experiential aspects of emotion. For instance, the typically short duration of surprise may be explained with reference to its formal object, the unexpected. For the length of time during which something can stay unexpected after it has been realised to also have actually occurred, is necessarily limited. Surprise at some particular specific object may be renewable, given inattention, memory lapses, or other ways of resetting expectations, but it is not durable. Though it may of course fade into more durable emotion-types such as bafflement or puzzlement. The formal object of surprise being the unexpected may also explain why surprise does not have a necessary hedonic valence, and need not have any hedonic valence, even
make them a viable object of fear. Being a courgette would usually not qualify, except in the case of some really unusual phobia, though being a spider might.

An object or target of an emotion is a **legitimate object** of that emotion insofar as the warrant for the emotion is that the object is held to instantiate the formal object at least in some vacuous way, such as appearing dangerous. Thus we have the minimal criterion of cognitive rationality for the emotion. The further evaluative criterion, as stated in the previous chapter, is that the warrant is held for a good reason. The more substantive evaluative criterion is that the warrant would still hold if the subject were in possession of full relevant knowledge. For instance, if I am angry because I think that someone has stolen my mobile phone, and it turns out that the phone is not in fact stolen, this would defeat the warrant for my anger.

There is also a further way in which warrant can be defeated: the cause (in Hume’s sense) of my emotional reaction may itself fail to qualify as making the target an instance of the formal object. For I might just have the wrong notion of the formal object. This is not always a straightforward matter to decide: the qualifications made to the cognition-claim above suggest that the formal object may display some significant degree of vagueness and polymorphism. Which is not to say that formal objects are altogether underdetermined. But I shall argue that some of the ways in which formal objects are determined, at the intersubjective and at the more idiosyncratically personal level, may also have a role to play in explaining this vagueness and polymorphism. First, though, I will outline briefly how the objectual claim features in assessments of the strategic rationality of emotion.

**Object-appraisal and the strategic rationality of emotion**

The assumption that emotion involves cognitive appraisal of an object also provides criteria for the strategic rationality of emotion. The features of the object that are salient to its instantiating the formal object are its **focal properties**: They direct our attention in appraising the object. De Sousa, (1987:335), specifies that "under certain circumstances, which define the standard case, the focal property is also the **motivating aspect** of the emotion". That something is considered dangerous,
then, would standardly figure both as the cognitive warrant for fear, and as the warrant for such motivations and action-tendencies as follow from fear. A typical fear-motivation would be the desire to get put as much distance as possible between oneself and the feared object.

The exact way in which an emotion-type motivates us may be as slippery a question as how we determine whether something counts as an instantiation of the formal object. But we are usually able to make broadly accurate predictions about how someone will react towards a target, based on what we know about their feeling towards it. The scared person will normally want to get away from the target. The angry person may want to stay and fight. Someone whose emotional reaction is one of happiness or enjoyment at contemplating the target may want to stay and bask. And again, further knowledge of the person who is the subject of the emotional reaction may help to get predictions right.

5) Beyond cognition and objects: Subjects, ideologies and emotion

The definition of the formal object, then, is where the intentional object and the cognitions involved in emotions meet. It is also crucial in deriving standards for the rationality of emotion. But where do we get these formal objects from?

The problems raised for voluntarism about emotion above suggest there is reason to think that the formal objects of emotion cannot altogether be a matter of choice. So it seems we do not get emotions, or their formal objects, by some Sartrean act of radical choice or self-definition. Some emotions, those that we seem to have in common with animals, may be largely determined by nature as to what elicits them, how they are expressed and what behaviours they lead to. But even for basic emotions such as fear and anger, there are social and individual variations with regard to all these features of the emotions. The same things do not make everyone angry, or angry in the same way. And at least some of these variations seem to be matters of choice, even for the basic emotions. I will explore that particular issue further in Chapter Five.
For this chapter, the last points I want to introduce are two further points or modification to the object-claim as outlined above. The first is that in humans, the formalisation of emotion by objects and grounds works through what de Sousa calls *ideologies of emotion*. While these are basically intersubjective, they also permit of a degree of particularity or even idiosyncrasy in the takes that agents have on them. And such particularity and idiosyncrasy may even itself be worked into the ideological formalisation. So the subjects of emotion - those that feel emotion - may be formalised in something like the same manner as the objects of emotion are.

But in formalising the subjects of emotion, ideologies reflect a central fact about emotions, and this is my second point here: That emotions reflect what those who feel them are like, as well as those things that emotions are felt towards. And that point, I shall argue in the next chapter, may crucially help to shift blame for problematic features of emotion off emotion as such. First, though, to return to the point raised at the start of this section: How do we come by the formal objects?

**Ideologies of emotion**

I will not speculate here about how this works for animals, though there might be interesting and informative points to learn from such considerations. In the case of humans, I think we can take it that the formal objects of particular emotions, and the ways we think about emotions generally, are not acquired by simple manifestations of hard-wiring, but rather through what de Sousa calls *ideologies of emotion*. The ideology of an emotion is "what we learn, in the process of being socialised, about (an) emotion: its moral and social significance, its place in the hierarchy of human states and capacities." (1980:278).

We can probably take it that what is "natural" in emotion will at least to some extent inform our ideologies of emotions. Of course, the naturalness of both the existence and the form of some emotions may be a moot point, given considerations of cultural variation: I will not deal with this just here. The point is that at least *some* elements in our ideologies of emotions may be just descriptive.

Mainly, though, ideologies of emotion seem to be normative, giving criteria of what count as intelligible and appropriate instances of an emotion. A child
acquires these ideologies, in the process of being socialised, by attending to how their parents describe name the child's own emotional states and those of other people, as those states manifest themselves in behaviour. And the content of this learning is not just descriptive, learning that what I feel now is anger, or fear - "Now you're just getting worked up and making yourself angry", says the parent, or "Now he's angry". It is also normative: "That's nothing to be angry about/afraid of"; "you should be ashamed of yourself"; "Now doesn't that make you happy?". De Sousa (1979) calls these contexts of learning paradigm scenarios. "Paradigm scenarios involve two aspects: first a paradigm situation providing the characteristic objects of the emotion (where objects can be of various sorts, sometimes more suitably labelled "target," or "occasion"), and second, a set of characteristic or "normal" responses to the situation." (1979: 145)

**Intersubjective and idiolectic takes on ideologies of emotion**

I do not want to push a particular account of emotional learning here. But there are three points I think are relevant regarding emotional ideologies and learning.

One is that ideologies of emotion, including concepts of formal objects, are acquired in social, intersubjective settings, and may reflect contingent features of those settings. This applies both to the influences of the individual's direct contacts, such as his significant others, and to broader social influences the agent is exposed to through education and acculturation.

Secondly, apart from the specific ideologies of emotions an agent is exposed to, more particular contingencies of the individual's formative influences may also be reflected in their take on the formal object.

The personal emotional ideologies of an adult human may for instance reflect accidental features of the particular paradigm scenarios the agent's emotional learning took place within. And their emotional reactions may reflect features of the deep objects of their emotions, those objects towards which the emotion was first significantly felt (see Baier, 1990). Who a person falls in love with, what they get angry about, what they are afraid of, and so on, may be shaped crucially by features of the deep objects that bear only accidentally on the emotion in question being *that*
emotion. It does not seem essential to love *qua* love, for instance, that its object should have this particular kind of personality or way of laughing, but the influence of the deep object may make it so. The deep object might for instance manifest its influence through one of Rorty’s magnetising interpretative disposition in a quasi-intentional set, or through the patterns of attention involved in the conceptual scheme of the emotion. Sometimes the magnetising disposition is informed by positive feelings towards the deep object, sometimes by negative.

And of course, the deep object that figures in the agent's emotional reactions is the agent's perception of the real-world entity in question. Jonah's mother, for instance, need not be accurately represented by the figure of resentment that influences Jonah's disposition to see his boss as overbearing.

The person themselves may not be aware, let alone reflectively aware, of the role these formative influences play in their emotional reactions. Though of course people are often aware of these influences, and may be quite explicit about them. People often have stories, for instance, of how their first love has influenced their subsequent choices of partners, or how some horrible incident involving an object has left them with a phobic reaction to it.

Thirdly: As well as deep objects and paradigm scenarios, the particularity of the agent himself may shape his emotional reactions, his patterns of choosing particular objects rather than others as targets for an emotion, his tendency to have particular kinds of emotional reactions rather than others. And this, too, might be part of the ideology, social and personal, of an emotion. We usually take it that particular kinds of emotional reactions reflect particular kinds of features in those that have the reactions. Such trait-ascriptions may be more or less relativised to particular kinds of objects. For instance, we might have different expectations about how a person will react to expressions of parental disapproval when he is a child, when he is adolescent, and when he is adult. And the description of the subject, too, may be more or less specific, from being based on mere kind traits and right down to particular personalities. For those that have closer interactions with an agent, predictions of emotional reactions may be relativised to a very high degree: A person's friends, for instance, may expect them to have very specific kinds of characteristic emotional reactions.
Formalisation by subject

Such expectations involve a formalisation of how we think of the subjects of emotions, those who have the emotions, that parallels the formalisation of the objects of emotion. These expectations, like those directed at objects of emotion, may partly be descriptive, based on what kind of agent demonstrably tends to be the subject of particular emotional reactions. But they are also partly, and sometimes heavily, constructed around normative assumptions. So we might say that a particular emotional reaction, like having a temper tantrum at the supermarket checkout, was not fitting for a grown man, and that feeling spite or envy is not appropriate for a friend towards a friend. And where the agent overall does not seem to fit the expected subject-description for their emotional reaction, we may ascribe the action to some sub-system of the agent, such as their "inner child", that the reaction fits better with.

Assessments of intelligibility, here as for the formal objects of emotions, may occupy an awkward position on the brink between descriptive assumptions and normative assumptions. And some assumptions that are actually normative may present as descriptive - we might think that this was the case for many assumptions based on race and gender, for instance.

Particular ways of linking emotional reactions to traits of agents may be questioned descriptively. And some (Harman 1999, Merrit 2000) have argued that trait ascription, more generally, involves ‘attribution error’ and cannot be used to predict attitudes and behaviours. Such attacks on the practice of trait-ascription might gain support from the apparent evidence of for instance the Milgram experiments on obedience (see Milgram, 1974), or Darley and Batson's "Samaritans" experiment (1973). In both cases, it might seem that the situation the agents were in, rather than anything about the agents, determined the behaviour of agents. And in both those experiments, the situationally determined behaviours were largely undesirable ones.27

---

27 The Milgram experiments, notoriously, involved subjects administering what they thought were potentially permanently damaging levels of electric shock to a fake subject, in the context of a supposed experiment on the effect of negative feedback on learning. Although the only pressure brought to bear was a repeated instruction by the "professor" overseeing the experiment to continue,
Subjects, objects and grounds in emotion

The way all this figures in the concerns of this thesis is as follows.

First, there seem to be cases where the object-claim will not hold, which raises basic problems for object-appraisal accounts of emotions as cognitive. Insofar as the target functions as the bearer of properties that provide a warrant for the emotion, whether or not we are able to identify a specific target seems crucial. And it also seems crucial that we should have attitudes towards this target that warrant the emotional reaction.

But it seems emotion may go awry on both counts. Some emotions, as already suggested, are rather like moods: we are not sure what it is we are feeling scared of, angry at, or happy about. And of course there are moods themselves, which, as argued earlier, may coalesce into object-directed emotions, or themselves be spill-overs from such emotion. If emotions involve representation of specific objects, moods are more like free-floating representations. And then there are blind emotions. Here, the subject can pick out who or what the emotion is about readily enough, but they may lack the kinds of warranting attitudes that would support their emotional reactions.

Now, these cases are by and large anomalous, and so we might not think them very important. We may also be able, based on what has been said above, to give accounts of blind emotions in terms of for instance the persistent subconscious influence of some more or less contingent feature of a paradigm scenario or deep object. Cases where the agent cannot immediately identify an object for their emotion may also be amenable to explanation in terms of the cognitions that could identify the object operating at a non-conscious level. We would normally expect that differentiation in behaviour or thought with respect to the object would show it

---

In the Darley and Batson experiment, a group of Princeton seminary students were asked to give presentations across campus, either on the parable of the good samaritan or on career prospects for theologians. They were then given directions that led them past a fake accident victim. The main factor in whether or not the students helped was whether they had been told to hurry to the presentation or not - their particular religious outlook, and what topic they had been asked to think about for their presentations, had little significant bearing on their actions.

That the object-claim typically plays a crucial role in grounding the rationality, and through that the ethical respectability, of emotions, is supported by considering how moods are often treated
up sooner or later. Mood may be thought to present more of a problem, but it is not clear that moods are fully unspecific as to object. They may at least sometimes be explained as affective states directed at some generalised object, as when they arise from considerations like: "the world is a nice place", "people are overall bad", or "everything I do goes wrong". While we might worry about what could count as evaluative warrants for the more extreme generalisations, moods so understood could still qualify as minimally rational. And where mood follows from a specific emotional reaction, such as upset after an argument with a loved one, it is at least minimally intelligible in terms of the central importance the loved one has within the agent's particular little subjective world.

The second point is this: Insofar as such cases can be explained without giving up on minimal rationality for emotion, it is largely by drawing on assumptions about the subjects of emotion as much as on assumptions about the objects of emotions. And this suggests that a crucial part of what determines emotional reactions - what they are, what they are directed at, how particular beliefs and desires figure in them - is what those are like who feel them. And that in turn suggests that much of what our actual emotional natures are like cannot simply be put to emotion's account. If we were different, our emotions might be different.

6) Emotion and the revised cognitive account

To recap: Emotion itself, on the account I have given here, is an affective mode of cognitive and motivational engagement with the world, involving but not reducible to desires, feelings and thoughts. It reflects its subjects as well as its objects, and it may be only weakly cognitive. And its cognitive nature is at least in part sui generis. And while some problems come with affectivity, it is no longer clearly obvious that emotion, so understood, is a problem per se. In the next chapter, then, I will examine how the kind of concerns that powered the objections to emotion might be blamed on sources other than emotion as such.

as somehow pathological by virtue of their lack of object-specificity - see for instance Hepburn (1984).
Emotion itself, I have argued, can be shown to be at least minimally rational. It has also been shown to be amenable to at least some degree of control and bringing into line with reason - though some worries were raised about reason. And some of the worries from Chapter Two about autonomy and objectivity and the ways they may clash seem to be back in play here. It is for instance far from clear what we should think about the proper relative role of the object and the subject in determining emotions. Worries about projection may pull one way and worries about choice in how one reacts to externals in another. And then again, there may be worries, if emotion reflects subjects, about whether it also reflects them as autonomous, or merely reinforces their heteronomous aspects by highlighting them with feeling. And the intersubjective element in ideologies of emotion may reduce worries about particular subjectivity at the cost of bringing back in worries about intersubjective heteronomy.

Insofar as we get emotion off the hook by putting the subjects of emotion on it, though, there may be worries on behalf of those subjects. It seems that some degree of internal coherence is important for an agent's flourishing, for instance through supporting a sense of self. And a degree of emotional internal consistency on the part of an agent is often thought to be important, if perhaps only instrumentally, for moral agency. The worry might then be that the subjects are too little reflected in their emotions and behaviour, and that it is situational features determine us emotionally, as Harman and Merritt argue. Of course, the kind of cases Harman and Merrit appeal to are ones where the agent behaves badly by acting on cues from situation rather than some more consistent moral character. But we could also note that some, even most, conceptions of objectivity suggest that being good moral agents also involves acting from situation not character - I will look into that issue again in the final chapter.
Chapter 4: external sources of problems with emotion

1) Introduction

I have argued for a view of emotion as an affective form of cognitive awareness - one that significantly involves feeling. It also has important links with and similarities to desires. The cognitive element in emotion emerged as something like a loose conceptual paradigm characteristic of the particular type of emotion in question. This account was offered with the qualification that the cognitive element might be merely weakly cognitive, proto-conceptual and quasi-intentional. For the purposes of this chapter, those qualifications can mostly be ignored, and I shall refer to the cognitive elements in emotions as conceptual paradigms.

Defined as an affective form of cognitive awareness, I argued, emotion in itself looks fairly blameless. And *qua* intentional awareness of objects, emotion still looks fairly blameless. However, I also argued that emotions reflect their subjects as well as their objects and causes. Or put less technically, emotions reflect what those who feel them are like as well as what the things that they are felt *about* are like. And this subject-reflecting tendency may affect individual agents' takes on the conceptual paradigms involved in emotions, as well as particular emotional reactions.

Whether the subject or the object features more strongly in the conceptual paradigm may vary, between agents, between cases, and between different emotional cultures. So may the way features of objects and features of subjects are linked to each other in the emotional paradigm. And the ideologies within which we do our emotional learning have their own part to play in determining these factors. I will have more to say about emotional ideologies specifically, and what limits there might be to their variety, in the next chapter. Given what has been said about them in the previous chapter, we may suspect that our emotional ideologies, too, may reflect what we are like as much as they reflect what emotions are like. For now, they will feature as just one more contingent feature that shapes subjects and their
perspective on the world, and hence the emotions of those subjects.

The argument I want to make here does not rely on emotion always reflecting what its subjects are like. In fact, I want to claim that we are also capable of emotional reactions that transcend contingent features of what we are like. Or at least, most of us are, and much of the time. But I do think that the kind of features that power the objections to emotion can be traced to some basic general features of human agents and to what could grandly be called the human situation. And I think that these features are not necessarily inherent to emotion qua affective mode of awareness. This is supported by the fact that our emotions change when we manage to transcend such contingencies. And emotion itself, as well as reason and imagination may have some part to play in achieving such transcendence.

Some ways the problems might be inherent to emotion

I will argue, then, that subjectivity, heteronomy and rationality-failure may have sources external to emotion. It can still be conceded that some problems are inherent to emotion qua affective mode of awareness. Specifically, the felt nature of emotion has an amplifying, vivid-making, effect on those features of the world that we find emotionally salient. What is highlighted in this way also gets added impact on motivation. And as argued before, feeling may sometimes be overwhelming. Both the general amplifying effect and the occasional overwhelming effect may affect our capacities for objectivity, autonomy and both cognitive and strategic rationality adversely.

But what we find emotionally salient may not entirely depend on emotion. In fact, determination of salience may to a great extent be external to emotion as such, and dependent on features of us and our situation. It may also be conceded that emotion has a partly constitutive role in shaping what we are like - formative emotional experience for instance has a crucial bearing on character. And since what we are like in turn tends to shape our situation, through shaping our attitudes and behaviours, emotion may play a part in shaping our situation, too. But the exact form that emotion's contribution takes still depends on factors external to emotion as such.
Probably, there are limits to how adjustable human emotion is: more on this in the next chapter. But such limits may say more about humans than they do about emotion.

Speaking of human agents, then: What are we like that have emotions, and what is our situation? We are many, each of us numerically and qualitatively distinct from other agents and from impersonal aspects of the world. To varying degrees, and with varying directness, we are also causally connected to this world and to other agents, and they and the world to us. These facts, I argue, are important sources of subjectivity, heteronomy and rationality-failure: I will discuss these in turn here. Dealing well with the problems arising from these features of us and our situation is also one of the most basic challenges we face as moral agents.\footnote{For accounts of morality that take these features of agents as constitutive of basic challenges for moral agency, see for instance Nagel (1986) and O’Neill (1996b), both of whom I shall return to in discussion below.}

2) External sources of the subjectivity of emotion

Epistemic fallibility

But the picture can also be expanded from the sketch just above. For one, we are epistemically fallible. We get the world wrong; we get other agents wrong, we even get ourselves wrong. The exact explanation for this a notoriously contentious issue, and I will not engage with it in depth here. I will discuss two more specific sources of epistemic distortion - perspectivity and individual particularity - further on. But it is also arguable - even without resort to indirect realism - that we are epistemically fallible in more general ways. Faced with the task of capturing reality, human understanding often and ultimately seems to fall short. Our senses may mislead us; so may our reasoning. Even at the most immediate levels, we may misperceive, and form the wrong beliefs. And that is (arguably) before we even start on interpretations or inferences, both of which we may also get wrong.

This epistemic fallibility is a fairly obvious source of bad subjectivity, or just of subjectivity \textit{tout court}. Now, if emotion is a form of higher-order evaluative
consciousness, applied to and informed by other cognitions such as beliefs, then this epistemic fallibility might be thought to matter to emotion only insofar as the raw material to which emotion was applied was itself flawed. The fault would then be inherited by emotion, and would not be with emotion itself.

If emotion in its more *sui generis* affective aspect is thought of as a form of first-order awareness, then perhaps similar worries apply to emotion. We might think that there were specifically emotion-salient facts out there in the world that we were failing to capture, and that when our emotions go wrong, they do so in a way parallel to that in which factual beliefs go wrong. This might suggest cause for worry about feeling, since I have argued for it to be at least quasi-cognitive. However, I do not think that feeling need be seen as a first-order form of awareness: it may supervene on other forms of cognitive awareness such as inchoate beliefs and desires. And when we say that feelings are appropriate or inappropriate, we need not be applying the kind of criteria we do when trying to match beliefs to some independently existing reality. The standards we aim for feelings to be appropriate to may be agent-relative. Such an agent-relative standard might be general, as the moral universalist insists it should be, or it may admit of a place for standards of appropriateness that are relative to particular agents.

We might also worry that emotion carries its own conviction, and that this is a source of epistemic distortion. Concerns about how it adds vividness arise here, as well as about its self-reinforcing tendencies in directing attention. Such tendencies would also be a problem for cognitive rationality. And possibly to strategic rationality, since such distorted perceptions might lead us to miss instrumentally important information. But here, again, the source of the problem seems to be the original misperception, though admittedly emotion features as an accessory after the fact. However, the fact that emotion directs attention may in and of itself be useful to us, so it may not tell decisively against it that it goes wrong in particular instances.

For instance, if we feel indignant that someone has been treated badly, it is not necessarily a bad thing that this disposes us to look for corroborating evidence. If

---

2 While I want to argue, as in the previous chapter, that emotion is *sui generis qua* mode of affective awareness, and is not reducible to beliefs or desires, I have no particular stake in emotion's being causally or ontologically independent of other mental kinds.
they genuinely have been treated badly, looking for evidence will be crucial in alerting us to exactly how, which in turn is clearly salient to working out how their grievance should be addressed. And if they have not been treated badly, we are more likely to discover this if we pursue the matter than if we do not. The worry, of course, is that our compassion might dispose us to see corroborating evidence where there is none, or to overemphasise what supports our view compared to what does not. And this is a genuine possibility. But it is a possibility mainly where the agent has general tendencies towards rationality-failures: the emotions of a more rational agent might follow the evidence more closely.

And it is not clear that such difference in background levels of rationality need be correlated with how 'emotional' the agent is. Some people with only very shallow feelings are still irrational in this way, and some people with strong feelings may be more rational. The basic issue seems to be how rational people are in treating appearances. A reasonably competent rational agent is aware that warrant, whether for belief or emotion, may be defeated. This may also be an issue of how invested people are in being vindicated in their attitudes, over and above concerns with getting things right. Of course, such investment in one's own rightness is usually emotional. But it is not clear that just having emotions condemns us to having this particular kind of emotional disposition.

These considerations in turn point to ways in which more particular features of agents may influence their moral capacities for better or worse. And that leads to more agent-specific issues of subjectivity.

**Perspective and relativity as sources of subjectivity - and a qualification**

General epistemic fallibility apart, we may also be epistemically handicapped by the way each of us experiences the world from within a particular perspective. The outlook of any particular individual is from inside their particular perspective - both in a literal physical sense, and in a more metaphorical epistemic and psychological sense. Perspective is bounded and constituted by the finite part that each of us is of the world. That I occupy - and in a sense am - a particular body, with particular kinds of sensory capacities, is a significant fact about me, and a
significant determinant of how I experience the world. So is the particular set of psychological traits, and the kinds of psychological processes, states and capacities that are mine. So is my physical location, and my social location, relative to other agents and relative to the world in general. Usually, then, perspective also brings with it relativity.

Emotion is often thought particularly liable to be influenced by perspective. How people and events seem to stand relative to us - how close or distant, and in what way related, if any - is of basic importance to our emotional reactions. Typically, we react more strongly to an event that is close up than to one at a further remove. This is moreover usually so with little regard for the relative sizes of the events in themselves. And this usually holds even where the person is willing to concede that the more remote event is in fact greater. Our own immediate problems usually have greater resonance for us than some objectively more significant world event. Though this is not always so: Sometimes at least we care more about the remote than the close by, and more about the genuinely bigger issue than the closer one. But the *ceteris paribus* greater vividness of the close and that which is immediately salient to us nevertheless exerts considerable force.

This worry again is not unique to emotion: it seems to come with perspectivity rather than emotion as such. My beliefs about my own current situation are usually likely to be more vivid than my beliefs about world politics. Emotion may however be thought to be particularly prone to perspectivity, given that emotion deals in vividness, and more specifically in highlighting by feeling. Perspective makes some things more vivid, based on how they stand relative to our perspective rather than any more general reason why they should stand out. To use one of Nagel's (1986) examples, that someone in the room has a headache is usually only particularly vivid if that person is me, at least as long as anyone else who suffers the headache does not make it an issue for me by making a fuss about it. Insofar as we feel from inside our particular perspectives, emotion might be thought to further amplify this perspectivity, and even to tend towards cementing it. These intuitions about emotion also leave a general temptation to ascribe the greater vividness of

---

3 See however Stocker (1997) for a defence of perspectivity in ethical reasoning.
some beliefs over others to differences in emotional salience. But this point need not work in emotion's disfavour, as long as the vividness is not misplaced. And if it is misplaced due to perspective, that need not be emotion's fault.

It should be noted that perspectivity itself need not be such a worry as long as the person is aware of it and endeavouring to correct for the ways it may distort perception. Where it becomes a serious concern is where the person believes the relativised picture of the world they receive from their particular perspective is in fact an accurate picture in some more absolute sense, or where they simply and unreflectingly operate as if this were so. Here, worries about perspectivity might combine with more general worries about epistemic fallibility. This is a particular worry insofar as any perspective may be assumed to carry with it its own blind spots as well as perspectival distortions. Perspective has a direct bearing on what we pay attention to, and patterns of attention to some things are also patterns of inattention to others. A further way of cashing out this notion of the blind spot is that occupying any particular internal perspective is to lose the perspectives external to it (cf. Nagel, 1986).

This last point is also compounded by the consideration that the internal perspective we each occupy is constituted by particular faculties for awareness of the world. This has a further relativising function, since the perceptual apparatus available to one individual or type of individual may not be quite the same as that available to another. The two may pick out different features of the world as salient. And where this is so, we may worry, for instance on Nagelian lines (1979h), about the loss of intersubjective intelligibility. Particularly so since intersubjective intelligibility plays a crucial regulating function in assessing the validity of perceptions. Which is of course not to say that such regulation is infallible: the many may be wrong as well as the one.

Again, these seem to be problems that arise independently of emotion, though again we might worry that emotion would tend to reinforce them. And at least one form the problem takes is specifically emotional: differences in emotional outlooks, or to use an older term, in temperament, is an important form of relativising difference in perspective. Given the motivational and cognitive impact
of emotion, generally speaking, this gives cause for specifically emotion-related Nagel-style worries. Differences in emotional outlooks, though, may be attributed to individual variations in character that need not be just due to emotion. It could also be attributed to variations in emotion-ideologies, blame for which may in turn largely be laid at the door of the relevant surrounding culture. So here again we might think that the source of the troublesome difference was not in emotion as such.

**First-person perspective**

Staying with worries about perspective for the moment, there is also a worry that goes particularly to the first-personal nature of perspective. For our own experiences are given to us seemingly direct. We do not have such access to the experience of others. As Nagel puts it (1986:20): “To understand that there are other people in the world as well as [oneself], one must be able to conceive of experiences of which one is not the subject: experiences that are not present to oneself. To do this it is necessary to have a general conception of subjects of experience and to place oneself under it as an instance”. Still, other subjects, and their experiences, are matters of conjecture to us in a way that we ourselves and our experiences are not. So the worry might be that the internal perspective inclines us not only towards egocentricity, but also towards some form of solipsism.

Two qualifications should be made right away here. First, I am not claiming some form of radical first person epistemic privilege. The difference in immediacy between knowing our own experience and knowing that of others may turn out to be one of degree only (cf. Fridlund and Duchaine 1996) - though even such a difference might still be a significant one. If my own experiences, emotions included, are

---

4 For an interesting if in this context somewhat off-topic re-examination of the problem of other minds, see Glendinning (1998).

5 Relatedly, I would also not claim first-person infallibility. Clearly, people can be wrong about what goes on in their own minds, and sometimes other people identify such supposedly “inner” goings-on better than the person who is their subject does. I tend to think that there are some aspects of emotions, such as their psychic and physical feel, that cannot be experienced except first-personally, though fully supporting this would involve a deeper examination of these phenomena and their role in emotion than falls within the remit of this thesis. But emotions can also be inferred from beliefs, desires, behaviours and public physical manifestations such as facial displays; others may do better than the subject themselves at identifying their emotional state from such cues.
significantly more vivid to me than those of others, we would also expect their motivational impact to be greater. And the incentive to recognise others as subjects in their own right, let alone give weight to their emotions, would be correspondingly lessened.

Secondly, I do not want to argue that a strong sense of oneself as a subject is inherent in our nature or situation: it may well not be, and as I shall argue further on, this point can be a source of heteronomy. Some people may not have a strong sense of themselves, whether in the thick or the thin sense (cf. Wollheim 1984, 1999), and may pay more attention to the emotional states of others than they do their own.

All I want to argue at this point is that one’s own experiences are present to one in a way that those of others are not, and that this difference tends to make a difference, for the subject, between self and other, self and non-self. But as the qualification just made above suggests, this kind of subjectivity need not be strongly present in all agents, though we might expect that it would be in most of them. We may feel pain at the suffering of others, but we do not feel their pain - or at least not as we would our own. This is not to say that some altruistic person may not suffer more deeply over the pain of others - but such cases are not standard.

We might expect that affect felt by subjects with egocentric and solipsistic tendencies would reflect such tendencies. And so we might expect affect to also reinforce the egocentric and solipsistic tendencies. On the point of affect reflecting the tendencies of the subject, however, it seems most of the blame may be shifted to the unfortunate constitution of the subject. We could also point out that emotion is here in the same boat with other modes of experience of any particular subject: Thought and perception, to, seem to be first-personal in much the same way.

**Embodiment and (first-person) perspective**

Both the egocentric tendency and the solipsistic tendency may be further reinforced by our perspectives also being embodied ones. The agent’s physical sensations are present to them in a way that those of other subjects are not. Embodiment is also a source of heteronomy, in that we are not physically invulnerable, nor physically self-sufficient. But that I and not others occupy a
particular body is also a significant and fundamental part of the way we distinguish between self and other. And that my body has a relation to me that other parts of the world, personal or non-personal, do not, is basic to a distinction between self and non-self. The degree to which such distinctions are salient, and in what way, may however be subject to considerable degrees of cultural and individual variation: Sense of personal space for instance, varies a great deal between cultures (Griffiths 1997). And the Lockean notion that self-ownership of one’s body is essential to one’s sense of self is also a relatively new one - or at least this is true of the universalised and "democratic" attribution of such self-ownership.

As far as embodiment being a source of subjectivity goes, the following may be significant. Embodiment has hedonic tone, distinguishable from but not unconnected to, that of affect. One’s own physical state may be a compelling object or trigger of particular kinds of affect, even where the physical state is not a part of being in a particular emotional state - as when someone suffering a headache becomes irritable. Like affect, again, physical sensation has powerful motivational and cognitive impact at the most basic experiential levels of agency.

In view of the hedonic tone of embodiment and its place in our sense of self, staying embodied and the qualitative experience of this embodiment will tend to be prime concerns to us. This holds even for those subjects who believe that their essential selves are beyond their bodies and will survive them. Given that embodiment has been linked to both egocentricity and solipsistic tendencies, this very natural tendency to be concerned with one’s own embodiment may however function as a handicap to moral agency on a number of levels. Physical cowardice, and our limited capacity for sustaining physical suffering, will for instance present limits on how much of a moral stand we feel capable of taking.

Moreover, physical feedback may be one or even the way in which we come to know our own affective states, though the scientific grounding of this view seems incomplete. Damasio (1994) notably argues for a view of affect as being continually fed back to the subject through their body - but this view has been present in the philosophical literature at least since James (1884). Though as stated before, not all emotions are straightforwardly identifiable in terms of physical sensations.
Specific features of human biology have also been used to explain features of our emotions - suggesting that these features of our emotions in turn may themselves be contingent. Thus the supposedly primitive nature of our emotions compared to higher cognitive functions has been explained with reference to the supposed primitivity of the limbic system, the amygdala, the so-called lizard parts of the triune brain (cf. MacLean, 1980). Such points have also been used to explain the apparent inertia of emotion as compared to higher thought. So have explanations in terms of hormonal biochemistry and the role of biofeedback from physical displays and behaviours. While not in a position to evaluate the evidence on these issues, they do intriguingly seem to suggest that emotions could be significantly different for beings with different kinds of bodies than ours.

This argument, by the way, need not suggest that, as time-honoured prejudice would have it, emotion is more brutally physical than "higher" mental kinds, more tied to the body. For one, other kinds of cognitive processing are radically affected by physical contingencies of human biology, such as the division of labour and the limited intercommunication between the two hemispheres of the cerebrum (see also Nagel, 1979c). But while emotion may not be uniquely affected by biology among human mental kinds, clearly the particular form emotion takes in humans may be affected by more or less contingent distinctive features of humans.

Perspectivity and intersubjectivity

Both the notion of perspectivity and the points about the first-person perspective and embodiment raise worries about failures of intersubjectivity. If we each perceive the world from within first-personal perspectives that are not just numerically but also possibly (and apparently actually) qualitatively distinct, the degree to which we can expect to share an experiential world is correspondingly limited. And since our sharing an experiential world seems crucial to morality this does raise real worries about our moral capacities. There may well be limits to how many emotional idiolects the kingdom of ends can accommodate before turning into Babel.

Clearly we do share an experiential world to some degree: However, our
ways of cashing out the understanding of this shared experiential world are achieved through the individual perspectives that we each possess, and these may impose idiosyncratic takes even on such experiences as we share. Taken too far, such idiosyncrasy may threaten to altogether undermine intersubjective intelligibility. Insofar as our individuality, numerical and qualitative, may be thought to have some value of its own, however, the demands for intersubjective “falling in line” may conversely be thought to pose a threat to autonomy. Then again some, even many, forms of idiosyncrasy may be due to influences outside the subject’s control: Formative influences or accidents of character or temperament may cause idiosyncratic takes on the paradigms for particular emotion-types, or idiosyncratic patterns of emotional reaction such as an overall tendency to aggression. I will come to the question of outside sources of heteronomy next, after one important qualification to these points about perspectivity.

Emotion and transcendence of perspective

What has been said about perspective above impugns perspective, not emotion. For it does not seem that the ability to have emotions is tied to occupying our own contingent perspective. We can have feelings on behalf of the occupants of points of view other than our own. We can for instance have emotions in empathy with emotions they are having. We can have sympathetic reactions to their situation, which need not depend on their having any particular emotional reaction to their situation. We may imagine ourselves in their shoes, with our own characteristic features, to see how we might feel. We may imagine ourselves in their shoes with their features, and if we cannot feel how they feel, we may still be able to imagine what they might feel. And we may have emotional reactions when picturing a

---

6 This way of drawing the distinction between sympathy and empathy is used by for instance Vetlesen (1994) and Goldie (1999). Sympathetic emotion might even be heightened where the object of sympathy is unaware of their plight. Part of the horror of the CCTV recording of toddler Jamie Bulger being led away from the shopping centre by the two boys that killed him is that Jamie himself is not aware of the danger, and appears to be trusting the two.

7 The difference between these two kinds of in-his-shoes-imagining is emphasised by Mackie 1977, in discussing different levels of universalisation. Mackie emphasises that as long as I put myself in the other person’s shoes with my preferences, my contingent features would still inform my moral judgements. NB also that putting myself in someone else’s shoes with their features, with a
situation *acentrally*, from no particular perspective within it (Wollheim, 1984). Or we might put together some more or less specific picture of an ideal moral judge, such as Smith's impartial benevolent observer or Aristotle's virtuous agent, and imagine what their emotional reaction might be.

And these are all capacities that we crucially draw on to inform moral agency. For instance, trying to stand back emotionally from one's own part in an argument, one might try to adopt the other party's point of view empathetically, or put oneself in his shoes with one's own and then his features. Or we can try to imagine, acentrally, how the argument would appear from some indefinite external perspective. And we might of course also try to imagine some specific third party's reaction, particularly if they are affected by the argument. And these various moves may be used together. This may raise the converse worry, as some writers, notably Stocker (1976a, 1976b), has argued, that morality does not leave room for the perspective the agent started out with, namely his own. And that brings us back to the next problem: Heteronomy.

3) External sources of heteronomy

**Coming to recognise oneself and others as subjects**

There are some reasons to think that solipsism is not quite so live a threat as some might think in looking at how we are constituted. Current psychology has moved away from the idea that the world of the infant is a blooming, buzzing, undifferentiated matter from which the infant only gradually learns to distinguish himself, and only later learns to distinguish other people. (We might think of that process as a move from pre-subjective to intersubjective heteronomy, with a brief stop for solipsistic tendencies in the middle). At the least, there is evidence that even at early stages, infants are aware of themselves and other people as distinct entities. Harris, (1996, cited in Goldie 1999), argues that infants as early as at five months

view to projecting what their emotional reaction might be, is distinct from empathising with some actual emotion of theirs, which I might or might not have been able to predict through in-their-shoes imagining.
recognise and respond to the emotions of others, being noticeably more still when
the adult in charge is angry, for instance. By twelve months, children were shown to
be capable of “social referencing”, identifying the target of the other person’s
emotion. So presumably there is, even at this stage, some recognition of the other as
a subject of intentional emotion-experiences in their own right.

Diminishing worries about subjectivity, however, has been shown on earlier
occasions to tend toward raising worries about heteronomy. A prime source of such
worries is intersubjective contexts. For objections to emotion from concern with
subjectivity, the usual worry is that such contexts seem to be under threat. For
objections from concern with autonomy, the concern is about what effect such
intersubjective contexts, when present, have on the agent.

Now, whether or not they are learned skills, the ability to recognise and draw
these distinctions between self and other, and between subjects and non-subjects,
seems vital to adequate moral agency, or indeed any kind of efficient agency. Failure
to recognise other people (or other animate creatures) as separate subjects, or to
recognise the boundaries between oneself and them and between subjects and
impersonal aspects of the world, is epistemic and conceptual failure at the most
basic level. And these are to some extent distinct skills: It seems quite possible
(however undesirable) that someone might master some but not all of them.

Moreover, recognising oneself as a distinct existing entity seems
fundamental to forming a sufficiently coherent entity in oneself to make an efficient
agent. It involves, for instance, coming to some sense of which aspects of the world
one is capable of controlling and how directly. Understanding this is vital not only to
efficient self-management, but also to managing interactions with the outside world,
other subjects included. And while such recognition of boundaries is not sufficient
for also respecting them, it would seem necessary for achieving this further skill.
Such recognition is a vital moral skill. For instance, it would crucially inform
participant reactive attitudes and their proper occasions. To treat a person like an
object is a form of moral failure, on Kantian accounts, and on liberal accounts at
least from Locke onwards. So is treating them as if they were merely an extension or
double of oneself, without recognising them as an independent subject in their own
right, with their own aims and their own point of view. O'Neill (1996b: Chapter Four) argues that both slavery and the kind of marriage laws that treated the wife as the husband's property are violations of this principle. And letting others treat one in these ways might perhaps also be thought a form of moral failure, if the notion of duties to self is thought to apply.

Now, as indicated in the previous section, even if the awareness of oneself as just one subject among others is not something we have to learn, there seems to be some built-in temptation to forget it as an inconvenient fact. And this might be particularly easy in the light of how much more vivid and real my interests seem to me, from my perspective, than those of other people do.

Nor would the recognition of oneself as one subject among others seem sufficient to ward off egocentricity. Some sense of equality between oneself and these others would seem necessary to avoid this threat, and that might not be achieved by the mere cessation of solipsism. In fact, the recognition of other people's existence, given egocentric tendencies, may not counteract egocentricity so much as present the egocentric person with an unwelcome sense of competition for goods.

One particular worry for autonomy is from how we react to being on the receiving end of such behaviour. And that worry feeds on two more general ones: one about our vulnerability and neediness, and one about a degree of heteronomy being inherent in intersubjectivity.

Vulnerability and dependency

In coming to terms with the existence of other subjects, of themselves as one among others, and of their own and others' place in a shared world, the individual also faces a number of challenges.

For one, as already indicated, there are issues of control. Recognition that one is separate from the world also involves a recognition of one's own limited powers. And insofar as we are causally and otherwise capable of being affected by the outside world, whether in the form of other subjects or in more impersonal ways, we are vulnerable. Depending on the outside world for sustenance, we are not self-
sufficient, and so are further vulnerable. Such vulnerability to externals is heteronomous, and such limited power as any one of us possesses seems insufficient to overcome this particular source of heteronomy. Given what has been said about the motivational impact of embodiment, above, we might also expect that this source of heteronomy would have considerable impact on how we are motivated, often in reactive ways, which might further undermine autonomy.

Secondly, given that we are unlikely to be capable of overcoming these sources of heteronomy on our own, and so remain vulnerable and less than self-sufficient, there is an implicit dependence on others. Depending on the degree of neediness and vulnerability of any particular individual, this might range from merely needing others not to actively harm one to badly needing their active help. Such need and vulnerability is likely to be present in at least some form and to some degree in any individual. So it would seem that this further form of vulnerability to externals (in this case other subjects) is also built into the basic human situation.

And if we are vulnerable to and dependent on others, their attitudes are thereby implicitly important to us. For we may assume that their attitudes will inform their behaviour - not necessarily in a straightforward or entirely dependable way, but still. This dependence may of course vary in degree and kind. The kinds of goods for which we are dependent on others vary enormously, and the nature and depth of our interactions with them vary accordingly. For instance, if need is limited to material goods, we may need at most superficial interactions: co-ordinated effort to secure such goods may require little personal depth of interaction.

But sociable animals, such as humans, also require goods in the non-commodity sense, such as affection and recognition, from others. And here the depth of interaction required to possess the goods is far greater. One might of course give more or less cynical or reductive accounts of why we should need the affection and recognition of others. Witness, for instance, evolutionary psychology-accounts, at least in their popularised and simplified versions, emphasising the adaptive benefit to the species of attachment-formation, and reducing the content of attachment to pursuit of a strategic goal.

However, the reductive nature of such accounts is not so much the point as
that they reinforce that we do exhibit such need and vulnerability. And, hence, that we are in some degree basically heteronomous. The evolutionary psychology account may, perhaps surprisingly, also tend to let emotion off the hook to a greater extent than a less cynical account. For if we came to be concerned with the attitudes of others through desires born out of need, then it seems desire rather than emotion might be to blame, at least at the outset. Emotion may now be involved in a rather richer manner in this form of heteronomy, but it is not (or at least not entirely) to be blamed for bringing it about. Of course, if we are concerned to make emotions that arise from concern with the attitudes of others look less self-seeking and more genuinely intersubjective, we might also want to argue against a reductive genetic account.

It is worth noting that emotion is also crucial to the formation of attitudes that allow interpersonal bonding at any level to be undertaken with a suitably diminished sense of the vulnerability of one or the other both parties. As well as sympathy and empathy, already mentioned, attitudes such as trust are fundamentally affect-involving (see Baier 1986, Jones 1996).

The intersubjective as source of heteronomy

The process by which we learn to understand our emotions, and ourselves, may also be significantly heteronomous. Emotional learning, I argued, happens through exposure to emotional ideologies and paradigm scenarios. In learning ideologies of emotion, the agent is being taught by intersubjective standards that may themselves be normatively informed constructs as much as descriptions. So there may also be ideological artefacts to his emotions and to his understanding of emotions. Such artefacts may be more or less peculiar to the ideological community in question. They may be more or less dubious in themselves - they might for instance discriminate against particular kinds of agents on grounds of ethnic background or sexuality.

And the agent may not have much, or any, choice as to which community he is part of. He may also have limited capacities for resistance and critical thought relative to his community - and the community may encourage such limitations and
even sanction attempts to overcome them. And not only are his emotions heteronomous in being shaped, and probably to a large extent induced, through emotion-ideologies over which he exercises little or no choice: as Wollheim (1999) argues, emotions, especially participant reactive emotions, may be more deeply heteronomous than that. For even where the agent has managed to rise above the contingencies of the particular emotional ideologies of his community, he may still be vulnerable \textit{in himself} to the approval or disapproval of others who are still using the ideology he has discarded. And in such cases we might think that a part of him was still attached to the old ideology. Burnyeat (1980) talks of Aristotle’s account of the flawed virtuous agent as one where there are "pockets of resistance" to virtue and reason in the soul of the agent, and that might be a useful picture here.

Intersubjective heteronomy may go beyond emotional learning specifically. A fundamental part of the recognition of oneself as a subject is being recognised as such by others. Coming to understand one’s own \textit{qualitative} distinctness will also be largely dependent on intersubjective feedback and comparison. And the means by which one represents oneself, the world, and others, to oneself, are (barring private languages) intersubjective in nature. (See again Fridlund and Duchaine, 1996, Wollheim 1999, Chapter Three.)

Insofar as an agent's sense of what he is, of his own standing in the world and relative to others is mediated by such intersubjective conceptions, it is also vulnerable to them. Neither the thick nor the thin sense of self is likely to be entirely autonomous, at least in its formative stages. We are not impervious to the ways in which others see us, though our degree and kind of vulnerability to the attitudes of others is varying. Individuals and cultures also vary greatly in the importance they attribute to the regard of others. So our very degree of autonomy may itself be subject to heteronomy, through the impact of contingent externals such as the nature of the social setting an agent is born into.

However general the intersubjective standards we apply to people usually are, the extent to which we go by the opinions of actual others or by some more transcendent standard varies. Some people might only take to heart criticisms made against them by some standard that they themselves share. Others, such as
Wollheim's *radically heteronomous* agents, might find the mere fact of being disapproved of deeply distressing regardless of whether they share in the views of those who disapprove of them. Insofar as we are, for whatever reason, sociable creatures, the threat of being made the object of negative attitudes, particularly by significant others, will tend to register.

Again, although emotion may reinforce this feature, and even play a constitutive role in it, merely being affective does not seem to be the issue: Affectivity taking the particular form of vulnerability to and need of the goodwill of others is. And as suggested, concerns with the attitudes of others may have roots in features other than emotion as such, namely our neediness and lack of self-sufficiency. Even if we do not accept such an account of how these emotions arise we may think that being emotional is not in itself to be other-directed, as in tending to be influenced by the beliefs and values of others rather than by one's own internal principles. For highly emotional people need not be other-directed.

**Heteronomy and significant others**

Another, related, source of heteronomy is to do with particular others. Coming to be aware of others beyond some thin general sense requires that one get a sense of their qualitative as well as their numerical distinctness from oneself. Deeper understanding of others seems to require, moreover, some effort at understanding how things appear from within that other, qualitatively distinct perspective. Aristotle argues in *The Nicomachean Ethics* that in friendship, the friend becomes "another self". Insofar as the friendship is motivated by affection for the friend in all their aspects, including that of being themselves a subject, we might expect that some degree of transcending one's usual perspective and taking on some of the perspective of the friend would be involved in forming the attachment.

Now, this transcendence need not involve giving up one's own perspective in order to identify with that of the friend. In fact, if the friendship is to remain a friendship between two distinct subjects, each needs to retain their own perspective, however much enriched by what they take on of the other's. However, there may be an inherent danger - in any personal interaction, not just friendship - that the point of
view of the one party may prevail unduly over that of the other, and even eclipse it, so that one person’s point of view no longer plays much role in the relation between the two subjects. When this happens, the other’s status as subject may itself be threatened: They are objectified.

This concern does not apply to cases where one party, convinced by the other’s reasons, modify their own point of view in a way that makes it more similar to the other’s: any healthy notion of autonomy presumably needs to allow for such a possibility.

Such imposition of heteronomy may moreover be something that one of the parties involved does to themselves, possibly without the encouragement or even the desire of the other person. Some people, faced with a situation of conflict, can see everyone’s point of view but their own, and consistently tend to under-assert themselves. The intersubjective balance may be a difficult one to strike, and we might want to recognise the loss of self as well as solipsism - heteronomy as well as negative subjectivity - as an inherent danger in the balancing act.

Now given that emotion plays a crucial part in interpersonal interaction, it may again amplify and reinforce whatever tendencies are already present. I have given some reasons, earlier, to think that these problems may have their sources outside emotion, in the way our perspectives are constituted, and in our social, needy and vulnerable natures. Moreover, given that emotion has a strong communicative aspect, it may be used manipulatively in interpersonal interaction, in order to undermine the autonomy of others. However, the possibility of abuse does not necessarily prove something to be inherently bad - though it does alert us to the need to be on the lookout for abusive uses.

**Sources of internal heteronomy**

One final form of heteronomy to be discussed here: when introducing the worry about heteronomy in emotion, concerns about internal heteronomy were also raised. Our emotions may arise out of aspects of ourselves over which we have little control. Character, over which we may have little choice or control, is a strong

---

8 For a fuller discussion of such issues in romantic relationships, see Friedman, (1998).
influence on emotion. Insofar as we assume that character and temperament are formed by external influences, the heteronomy too will be external. But we might think that people do not start out with a tabula rasa, personality-wise. However, that suggests our characters may just be givens, possibly of a genetic nature, which only serves to move the source of heteronomy further inwards. In this respect, character would resemble the biological givens (or supposed givens) of the particular ways in which we are embodied, and the particular obstacles to autonomy those might be thought to bring: character may at least initially be as little a matter of choice as bone structure.

Here, again, the problem does not seem to be emotion as such. Unless we assume that subjects do start out with clean slate and options of radical choice as far as any qualitative distinctness goes, we will have to concede that to a large extent we stand in much the same relation to ourselves as we do to other people and to impersonal aspects of the world. In other words, faced with something not altogether of our choice or under our control, and many of whose features we may wish different. Which is not to say that we are altogether powerless in the face of our own natures, but some lack of choice, control or power to change may have to be conceded. Again, emotion may be a factor in reinforcing what is there already, but it seems that we can identify other sources of internal heteronomy. And where we are capable of transcending these features of ourselves, we can expect emotions to follow. In some cases, emotions may even lead. For emotions might not only transcend egocentricity, but also heteronomy: self-assertive emotions, as well as other-directed emotions, may have a crucial part to play in moral agency.

4) External sources of rationality-failure

And that brings us to the final heading under which I raised worries about emotion in Chapter Two: Rationality. I shall have relatively less to say under this heading here, as much of the concern with rationality in emotion has been addressed in discussion of the cognitive theories in the previous chapter. That discussion was however largely concerned with trying to find whether there were intrinsic features
of emotion powering the worries: the concern here is with outside sources of the features that power the normative objections.

For the main part, outside sources of rationality-problems seem to me to be the same ones that raise problems for objectivity and autonomy.

For instance, as indicated in the previous chapter, insofar as emotion is seen as a higher-order form of consciousness, it is liable to inherit flaws from lower-order forms of consciousness on which it draws. And, given our epistemic fallibility, the possibility of such flaws occurring seems built in. For reasons already cited, emotion may be particularly prone to compound such problems, by making what has gone wrong also go vivid. But again, it is not clear that it can be entirely blamed for them. Also, as I have argued earlier, other faculties such as reason may themselves be flawed, and insofar as emotion is guided by them, it will share their flaws. Of course, emotion might be thought to be all too little guided by reason. But if reason too can be impugned, there does not seem to be clear reason to think emotion necessarily worse off.

As indicated above, that each individual is constituted, at the ontological level, as a distinct existing entity with a particular location (literal and metaphorical) within the world, is reflected epistemologically in the perspectival and first-personal outlook of the individual on the world. But the constitution of that epistemological outlook has been argued to have an inherent solipsistic tendency, which will tend to obscure to the subject the salient fact about the world that it contains other subjects.

Now this tendency to solipsism is worrying from the point of view of cognitive rationality, since it not only distorts the facts but may also tend towards bad faith. It may also be strategically irrational, insofar as losing sight of actual features of the situation may be thought to make successful management of it less likely. However, the source of these distortions of rationality does not lie in emotion, but in the constitution of subjects.

Similar points apply to the distorting effect of perspective, though this may be rather more complicated matter. Insofar as it affects correct perception of the actual situation, if only by what it leaves out, perspective may be detrimental to cognitive rationality. And forgetting to correct for perspective where one is aware
that it applies and that it may distort seems to be a blatant case of cognitive irrationality. For strategic rationality, the effect is not so clear: Where sticking to a perspectival outlook is to the agent's strategic advantage, it may not be too worrying, but such an approach does seem morally questionable as well as likely to backfire.

However, given that perspective seems to come with the constitution of the agent qua individual agent, not using the information perspectival information gives us about the agent’s nature would in itself seem irrational. Cognitively irrational, since it means ignoring some relevant, if very much relative, and far from disinterestedly perceived, truths. And strategically irrational - since we might miss useful clues as to how best to promote the interests of the agent, which we will need to know even if we believe his interests to be defeasible in particular cases by more weighty concerns.

Similar arguments also apply, mutatis mutandis, to other contingent features of the agent and the way they affect the agent’s outlook. Consider the distorting effect we might expect from internal heteronomy. Now the features of the agent, as well as the features of emotion, that underlie internal heteronomy, tend towards distorting the agent’s outlook on the world. In that respect, they would seem to involve rationality-failures. On the other hand, as I have argued in the previous chapter, where an emotion does not make rational sense in the way that it picks out or construes its objects, it may make sense in terms of the nature of the subject who feels the emotion. Thus it may give us crucial information about the subject. While the subject’s outlook may not be more than at most minimally rational in itself, engaging with it seems to make rational sense insofar as it affords us important information about the subject. We may moreover crucially need to build on such information about the subject and their distorted state if we are to be able to move them on from this state.

Some forms of rationality-failure typically attributed to emotion might also be due to the complications any subject faces in dealing with their own contingent natures, their relation to the outside world, and their interactions with other agents. This all the more so in that each agent is a complex entity, with a multiplicity of features and roles (sometimes apparently contradictory ones) that affect how they
engage with other similarly complex subjects. Simply given the complicated balancing act involved in giving headroom to all these features of our situation, let alone doing justice to others, and ourselves, local breakdowns of rationality, cognitive or strategic, can hardly be unexpected.

Emotion may play a part in such breakdowns insofar as its paradigms involve both subjects and objects, and relate both to grounds - which suggests that emotion in itself is highly complex. And given its three main variables - objects, subjects and grounds - it is liable to failure on all three counts in individual cases. So someone might get the conceptual paradigm wrong, or the object, or be mistaken about themselves and their role in the situation, and their emotions be badly warped as a result. So the person who sees themselves as the poor suffering victim of others' wrong-doing may as a result develop tendencies to self-pity, self-righteousness and distrust of others, adding this to the damage that may or may not initially have been done to them. But insofar as I have argued for external sources for all these kinds of errors, emotion itself still does not appear to be the problem. We might get the wrong idea of the object or subject of emotion through epistemic fallibility, independently of emotion. And there might be distortions to our conceptual paradigms for emotions through idiosyncrasies and through distorted social ideologies of emotion.

5) Contingencies of outside sources and contingencies of emotion

Some possible ways to correct for bad contingencies - and possible roles for emotion in these

To the extent that we emerge from the account above as morally handicapped, it is perhaps because we are. Which is perhaps also exactly why we need morality. (Apart from to deal ethically with problems of natural scarcity, that is). However, the account above suggests that some of the same features that cause problems may also have positive roles to play. For instance, our separateness might be thought to be if not a necessary, then certainly a major contributing, element of
our capacity for autonomous agency. Similarly, the intersubjective dimension that in some respects may constrain or undermine this autonomy also provides a way of overcoming the impact of forms of negative subjectivity, since it allows us a crucial means of transcending (and expressing) our own particular outlook and interests.

An anti-utilitarian might also think our vested interest in our own particular perspective and welfare provided a welcome bulwark against temptations to add across persons\(^9\) - which is not to say that such vested interest does not also have inherently normatively undesirable sides to it. The intersubjective dimension also provides a counterweight to our seemingly built-in solipsist tendencies. If moral agency emerges from this as a complicated and even inherently unstable balancing act, it might just be that that is because it is.

Now if emotion could be different if we were different, then to the extent that we can transcend the limitations of the way we are constituted emotion might reinforce this. For instance, we are (mostly, if not all and always) capable of transcending our particular perspectives, whether towards some more general point of view or towards the points of view of particular others. Insofar as we are capable of assuming these other points of view, we can also be subjects of emotions that reflect these points of view, and not just our original internal perspective. And we are in fact capable, notwithstanding the tug of egocentricity, of assuming such perspectives. Insofar as this is the case and such capacity for transcendence is a help to moral agency, emotion too will have a positive role to play in overcoming our built-in limitations. And since desired emotions can to some degree be induced, for instance by acting as-if, it seems emotions might not just follow but lead. (See also Apostoleris and Laird, 1996). And since emotion gives us information about the subjects of emotion, paying attention to emotion, our own and others', may provide us with information vital to moral agency.

The considerations raised here about the nature and situation of agents also point towards the different kinds of categories of participant reactive attitudes Strawson operates with. Insofar as we ourselves may be on the receiving end of the actions and attitudes of others, and are vulnerable to them, we would expect to have

\(^9\) For some suggestions that agent-relativity of this kind might have positive roles to play in
reactive attitudes on our own behalf. Insofar as we are aware of ourselves as being in turn capable of affecting others, to see the similarity of this situation to one in which they affect us, and to be concerned on their part, we would expect to have self-reactive attitudes. And insofar as we recognise that other agents affect each other, and take this to be a matter of concern, we would expect to have vicarious attitudes.

But what makes the difference here is not whether we have emotions or not as such, but how we are constituted who have the emotions. And this again suggests the McGuffin point: What is really at issue is not emotion, or not just emotion, but how these various features of agents and their situations that create problems feature in morality. I will look at this issue more closely in the final chapter.

So emotion is not the issue as such: mostly, we are. If we were not capable of extending sympathies to other selves, we might expect to have only self-reactive attitudes, and to react to only as to agents and not as to moral claim-holders. It is the possession of emotional paradigms that take other subjects as objects in this particular way, and on grounds that give them the status of moral claim-holders, that is crucial, not simply the possession of the capacity for affect itself. The capacity for affect itself might take entirely self-centred forms (and sometimes actually does).

In the next chapter, I will look more closely at how some of the issues raised about emotion-ideologies and conceptual paradigms in emotion might feature in a more positive account of emotion in moral agency. I will also pick up again some of the basic issues about emotion in moral agency that introduced the thesis. For now, I should just like to add a coda that might defend some of the more particularly human aspects of our emotions.

**Human emotion**

Most of the worries raised about subjectivity here are about the partial and flawed ways in which we are *motivated*, rather than focused on the epistemic handicaps as such. Interestingly, some significant aspects of our emotional lives may in fact be shaped by awareness of the epistemic limits imposed by our particular perspectives. And arguably by awareness of the fallibility of our beliefs, too.

---

Most of our emotions, as Gordon (1987) argues, are *factive*: Directed at things that we take to be actually the case, or at least insofar as we picture them, however hypothetically, to be the case. But some are what he calls *epistemic*: Directed at states about which we do not have certain knowledge, either because they are in the future, or because we are in some other way blocked off from certain knowledge about them. Notably, hope and fear are typically epistemic.

The sense of uncertainty involved in epistemic emotions may be largely subjective, and sometimes verge on the irrational. So a parent might feel anxious about the welfare of their child even when they have only just talked to them on the phone and have every reason to think that they are home safe. (Sherman, 1997) Or someone might compulsively keep checking that they still have their tickets and passport on their way to the airport, despite there being no likely reason to think that these items have been left behind. Or someone might continually demand evidence of affection from their partner, and watch vigilantly for signs of what the partner is thinking and feeling, because the terror of not knowing the partner’s state of mind gnaws at them. And although their cognitive and strategic rationality is questionable, such cases are also intelligible. For we can understand that there are concerns that would motivate us to seek certainty and to feel pained by its continuing to elude us.

These problems apply to epistemic emotions insofar as they are forward-looking. Some similar problems apply to desire, which is also typically at a temporal or epistemic remove from its satisfaction. While we might want something knowing we have it, usually felt desire is for what we are not sure of getting. As de Sousa argues (1987: Chapter Eight), this raises some problems both about knowledge and time in relation to desire. And since desires feature in factive emotions, they too might have some of this temporal and epistemic dislocation.

To see that these parts of our emotional lives may be contingent, we might appeal to a fictional example of a being with a rather different emotional structure than ours. In Wim Wenders’ 1987 film “Wings of desire”, the angel Damiel, having fallen in love with a mortal woman, opts to become mortal himself. Love is not all his motivation, however. For he has also had a yearning for a while to have the emotional life of a human. As an angel, he explains, the certainty that he is living in
the best of all possible worlds is there for him all the time, dampening down emotional reactions. And while not omniscient, he at least sees events *sub specie aeternitatis*, without the level of uncertainty that faces humans. As a guardian angel, he can also read the minds of humans. And his body is in a sense only notional: The world has no causal impact on it. But he envies humans precisely their being in time, their incomplete knowledge, their physicality, and the intensity these things give their experience. He too wants to be able to “say ‘oh’ and ‘ah’ and ‘hey’ and not only ‘yes and amen’”. He wants a life with some room for whys and maybes. And perhaps this might suggest to us too that there is something good about our specifically human emotions.

6) Emotion and external trouble-sources

So can emotion be exculpated? Well, yes and no. Some of the distinctive features of emotion have previously been shown to be ones that validate concerns. Emotion tends to amplify worrying features anyway present, which shifts some of the blame, but hardly leaves emotion itself looking truly innocent. And emotion seems to be involved in at least part-constitutive ways with several of the kinds of features that blame has been shifted onto, which makes it look like an accomplice both before and after the fact.

However, for most of the kind of worries raised about emotion in Chapter Two, there are compelling outside sources, which may take if not all then much of the blame for what is usually laid at emotion's door. And this suggests that emotion as such is not the core problem. Granted, shifting blame to external sources will not get rid of brute feeling-objections to emotion, but those can be argued against on other grounds, as I have in Chapter Three.

Insofar as the argument above exculpates emotion, however, it conversely inculpates the subjects of emotion. We emerge from the account given in this chapter as morally handicapped in fundamental ways. As I have argued above,

---

10 Though contrary to expectation, Wenders' angels are rather ordinary-looking and clearly adult, not young and androgynously beautiful: one of them is played by Peter Falk of *Colombo*.
emotion may not be doing much more than reflecting features that are there anyway. Though given the internal role such features of our natures and situation play in setting the paradigms of particular emotion-types, this is presumably what we could expect. This does however raise the possibility that emotion might be different, and less problematic, if our grasp of reality were firmer, our motivations and perceptions more appropriate, our natures more transparent to us and more under our control. Or if we were less shut off from each other, if we were more self-sufficient, less self-serving - or both - and so on. On the other hand, the entities whose new improved emotions these would be would be rather unlike human agents as we know them. And by that token, the kind of moral problems they faced would not be the same ones that we face.
Chapter 5: Re-evaluating emotion in moral agency

1) Introduction

I have argued, in Chapter Three, for an account of emotion as a form of affective cognitive awareness, and that such an account may go far towards meeting normative objections to emotion playing a part in moral agency. In Chapter Four, I also argued that some of the features that power normative objections to emotion have sources external to emotion as such. Rather, the sources of the problems may lie in certain aspects of human nature and the human situation, which are then reflected in our emotions. Emotion, understood as an affective mode of awareness with important links and similarities to desire, seems fairly blameless in itself. In fact, as I argued in Chapter One, emotion, so understood, has a necessary and irreplaceable role to play in informing and motivating moral agency.

Insofar as this exculpates emotion, though, it seems to inculpate us. But I have also argued, in the previous chapter, that we have capacities for transcending those features of our natures and situation that give rise to the problems. Insofar as we are capable of such transcendence, our emotions will tend to change accordingly. And emotion may also have an important role of its own to play in achieving such transcendence. So there are grounds for optimism regarding both emotion and human nature when it comes to transcending the problems - even if there are also reasons to qualify that optimism.

In this chapter, I will examine further how emotion might be adjusted and changed, and how this bears on issues of emotion’s role in moral agency. Specifically, I will look at how what (following de Sousa) I have called ideologies of emotion play a crucial role in shaping and even transforming emotion. Ideologies of emotion, I argue, crucially exemplify human capacities for creative self-determination. Because they shape and change emotions, ideologies of emotion also play a crucial role in shaping us: as argued earlier, our emotions shape as well as being shaped by what we are like. Ideologies of emotion also play a significant role on both the descriptive and the normative sides of theories of moral agency: They inform both what is thought possible and what is thought desirable when it comes to
our emotions. But what ideologies of emotion present as desirable may of course also itself be crucially informed by moral theories.

The possibility of ideologically informed adjustment of emotion also suggests a further way in which the objections to emotion might be met. For as creatures with reason, we may be able to use ideologies to bring emotion further within the realm of rationality. And if we can transform emotion, it seems we can also control it - or at least, we can gain as much control as our capacity for transforming emotion will allow. And again, if emotions can be adjusted to fit some more accurate mode of perception, we might be able to make them more objective. Assuming, of course, that these are all worthwhile aims - I have already suggested (in Chapter Two) qualifications on how far all three aims are realisable and worthwhile.

I have mentioned earlier a number of other ways that emotion might be controlled or changed, albeit often only in a limited way. So we may adjust emotion by argument, appealing to the cognitive element in emotion. Or we may use acting as-if to adjust emotion through behavioural means. We may deliberately make use of physiological means like biofeedback. We may change how we feel by changing our point of view - and so on. Any of these methods, I take it, could be harnessed to bring emotion more in line with our ideological preferences.

So, too, we might hope that the transformative power of these ideologies could be used to improve moral agency, and to guide us when using less overtly ideological means to this end. But I will also look at what limits there are to such transcendence, and what implications such limits have for issues of emotion's place in moral agency. For we might have legitimate misgivings about the idea that emotion is infinitely adjustable and can be brought around to whichever approach to morality we happen to favour.
2) Changing emotion - and limits on change

Limits on the possibility of change

As stated above, one reason for such misgivings is that it seems likely there will be limits on what kinds of transformations of emotion are possible. Emotion qua affective mode of awareness might theoretically take any form. But, as I argued in the previous chapter, the emotions of particular kinds of entities in particular situations would probably take rather more particular forms: especially so since emotions reflect their subjects and objects. As I argued in the previous chapter, the way we are constituted as agents - separate, but interconnected and interdependent - significantly shapes our emotional natures. So, too, does the particular way in which we are embodied - our species-specific psychobiology. As far as such counterfactuals make sense, we could imagine that agents who were differently constituted than we are would have emotions, or emotion-analogues, rather different than ours.

Of course, emotions may still surprise us even where we are reasonably well informed about their subjects and objects. But I have argued (Chapter Three) that in the light of such information, emotions are usually at least intelligible after the fact, even where they are not predictable before the fact. And while human nature is diverse and plastic, and may to some degree be a matter of choice, it does not seem plausible that it can be shaped just any which way. So what the subjects of emotion

---

1 I will not here go in any great depth into the issue of whether affective modes of awareness significantly different than ours would really be emotions. For a closer discussion of the notion of emotion as a natural kind, and the degree to which our concept of emotion is necessarily empirical and contingent, see e.g. Griffiths (1997), who discusses implications for emotion-research of Kripkean approaches to natural kinds. Griffiths suggests that we can only be fairly confident of the existence of emotion in creatures whose neurophysiology resembles ours - so it would be a moot point, for instance, whether invertebrates such as octopi have emotions.

For my purposes, the "essence" of emotionality is that it involves affective awareness: There seems to be no reason in principle why this should not take forms - and have physiological or other bases - rather different from ours. For instance, there does not seem to be an obvious logical reason why emotions should necessarily tend towards inertia. Though I have suggested that such inertia may have some practical advantages: Where too-nimble reasoning may go wrong, it may be useful for there to be resistance from other forms of awareness. And physical or biological necessity might turn out to underpin what seem to be contingent features of emotion. In the absence of conclusive evidence or arguments for this, though, I will take it here that a number of features often thought of as inherent to emotion may be due to features of human nature and the human situation - as I argued in
are like, and what their situation is like, will significantly shape and limit what flexibility emotion is capable of—considerable though that flexibility may be. What we are able to say about the nature and situation of humans may of course be highly abstract and general—but it need not therefore be insignificant.

There may also be more emotion-specific reasons for such limits on flexibility. As suggested earlier, the distinction between what is inherent to emotion and what is due to external sources may be a fine one—but I will try to keep it as clear as possible in what follows.

**Limits on the desirability of change**

A further reason for misgivings about ideological transformation of emotion is that there may be limits on what changes to emotion and human nature are desirable. The ways in which emotions can be shaped need not all be good ones. At the risk of being too optimistic about the coincidence of the good and the right, it seems to me that bad transformations are unlikely to be compatible with flourishing, at least in the longer term. And hampering flourishing might significantly limit the appeal of bad transformations—though again perhaps only in the long term. Or so we would at least hope. And insofar as a significant aim and justification of morality is the promotion of human flourishing, this lack of emotional flexibility may serve a useful purpose. For limits on emotional flexibility might help to restrict the range of long-term practicable moralities to ones compatible with promotion of human flourishing. Beyond this restricting function, emotion also plays a more positive role in alerting us, through patterns of emotional salience, to what will promote such flourishing.

A related point, mentioned above, is that I have already queried the validity of the assumptions supporting the objections to emotion. Perhaps deliberate use of the transformative capacities of ideology could make us more rational, objective and autonomous. But this might not be quite so much of an improvement as hoped for. And given that objectivity, rationality and autonomy may not be jointly or severally coherent aims, we may also want to exercise some caution as to how and how far we
pursue them.

Also, I argued in the previous chapter, problems with emotion are largely due to our nature and situation rather than to emotion as such. And if that is so, we might also worry that those features of us and our situation that cause the problems will be detrimental influences on our ideologies of emotion. In other words, the ideologies by means of which we attempt to improve emotion may themselves exhibit our subjectivity, heteronomy and failures of rationality. And for the same reasons, the normative standards involved in our ideologies may themselves be flawed. And the same goes for our moral theories. And this in turn points to the recurring theme of emotion as a McGuffin. It seems emotion as such that may not be the problem. Rather, what is at stake is unresolved issues in our thinking about moral agency. This is not to argue for some kind of defeatism about moral progress - but it is to argue the need for care, caution and self-critical awareness on the part of those attempting moral improvement.

In making these arguments, I do not want to defend some unbridled psychological naturalism about what constitutes good moral agency. Clearly, givens of what we are like and what we feel do not yield indefeasible sources of value. And clearly there may be room for improvement in human nature, both on a species-level and as it manifests itself in individual agents. Nor do I want to argue for some rigid notion of what human nature is and can be. Our specifically human nature, including our capacities for moral agency, is significantly second nature, a result of learning and of our capacity for self-determination, as individuals and as a (sociable) species. So human nature is not set in stone. But I want to resist the idea that the standards by which such adjustments and corrections are informed can entirely transcend the givens of what we are like as individuals and as a species, or that it should. Rather, the givens may significantly inform and even constrain the regulating standards. Our self-regulation, so understood, takes place within the context of human desires and goals and emotions, not from some entirely detached vantage point outside it. And becoming better humans seems a more worthwhile - and a more realistic - goal than becoming something altogether superhuman or beyond-human.  

\[^{2}\] This is of course broadly an Aristotelian argument, as long as we emphasise the this-worldly
3) Ideologies of emotion and ideologies of emotionality

To start by recapping what ideologies of emotion are, and their significance: An ideology of emotion is, as stated in Chapter Three, "what we learn, in the process of being socialised, about (an) emotion: its moral and social significance, its place in the hierarchy of human states and capacities" (de Sousa 1980:278). I shall have more to say about the elements that make up such ideologies further down. But some further clarification of the basic concept itself may be in order here - also because my use of it may not exactly match what de Sousa intended. Ideologies of emotionality may take a bewildering number of forms, even on de Sousa's account, and come at us from a number of sources: From paradigm scenarios, from personal and institutional interactions, from popular culture and high culture.\(^3\)

First, de Sousa's definition suggests there may be ideologies both of particular emotion-types (anger, fear, gratitude, and so on), and of emotions as a class. For the sake of avoiding ambiguity, I shall refer to ideologies that go to emotions as a class as ideologies of emotionality. Among these, I shall distinguish between general and particular ideologies of emotionality.

A general ideology of emotionality need not make distinctions on a normative or other basis between particular types of emotion: it is directed at emotions qua emotions. Lower-case stoicism, for instance, is a (negative) general theory of emotionality. Blanket contrasts of emotion with reason or other mental kinds also rely on general rather than particular ideologies of emotionality. A general ideology of emotionality may for instance delineate what areas of agency aspect of Aristotle's ethics, rather than his emphasis on reason as the pinnacle of human achievement. Aristotle's account however emphasises, in *The Nicomachean Ethics* as well as in *De Anima*, that emotion is a faculty we have in common with animals. The credit for the greater flourishing of emotion humans are capable of goes, for Aristotle, to reason and correct habituation.\(^3\)

On the point about culture, it is significant that the revival of interest in emotion has to a great extent been focused, both in moral philosophy and in aesthetics, around fictions: novels and dramas in particular. I shall not pursue this theme in depth here, but some of those that have taken inspiration from literature have already been mentioned: Elster, Hepburn, Nussbaum. A great deal of work on the philosophy of mind aspects of emotions also goes on in the context of broader writing on emotion in the arts, particularly in music, with writers like Budd (1985), Madell (1996), Ridley (1995) and Matravers (1998).
may properly be influenced by emotion, and which may not. Thus, for instance, our culture generally holds that certain personal areas of choice, such as choices of friends or sexual partners, may legitimately be determined largely or wholly on the basis of feeling. There may not even be much in the way of limits imposed on how random or contingent the feeling in question can be and still legitimately determine the agent’s choices. “I just like him/her” is in most cases a perfectly adequate reason for being friends with someone, as just not liking them is for not being friends with them. In other areas, such as those governed by law, ideology dictates that feeling, if it figures at all, may do so only within strictly delineated formal constraints. Generally, affective states - moods and desires as well as emotions - are permitted to influence behaviours and attitudes in the private sphere in ways that would be much more restricted in the public sphere.

A particular ideology of emotionality, by contrast, will have a theory - or several - of how different emotion-types stand in relation to each other and to other states and capacities. For instance, the view that some emotion-types are more basic than others, or that some emotions are natural and others not, would be part of a particular ideology of emotionality. Particular ideologies of emotionality also determine what different types of emotion are recognised to exist - a point on which there is considerable cultural variation. In normative respects, a particular ideology of emotionality may rank some emotions as good (e.g. friendly feeling, sense of justice), and others as bad (e.g. malice, envy). It may hold that some emotions, such as those involved in attachments to particular others, take precedence over some of the deliverances of impersonal reason, while others, such as a craving for self-centred satisfactions, do not. A particular ideology of emotionality may also distinguish between emotions that have a legitimate or necessary role in moral agency (“moral emotions”) and emotions that do not.

While these distinctions will be important here, in what follows, I will tend for simplicity to refer to ideologies of emotion and of emotionality together as just ideologies, unless there is a need to make it very specifically clear which kind is at stake.

Secondly, the kind of ideologies (of emotions and of emotionality) that I
have in mind here are not just the views held by the culture in which an agent is
socialised, or by significant others encountered in the process of socialisation -
parents, teachers, peers. We can also add, as I have suggested in Chapter Three, the
more individual ideologies that particular agents hold, by choice or otherwise. "Or
otherwise" since such ideological *idiolects* may be significantly heteronomous. As I
argued in Chapter Two, idiosyncrasy is not the same as autonomy. Ideologies may
be influenced by contingent environmental influences and by quirks of individuality
over whose appearance and formation the agent had little or no control. Someone
who is naturally inclined to be determined and confident may for instance tend
towards a strongly negative or dismissive attitude to emotions like fear and
uncertainty, and towards emotional ambivalence. But where this attitude arises out
of character-traits that the person just *happens* to have, it hardly qualifies as
autonomous.

Similarly, idiolects formed in reactive attempts at asserting autonomy may
by that very token be significantly heteronomous - this would probably apply to
most forms of adolescent rebellion. But we can also allow that idiolectic takes on
ideologies can be autonomous. The main point here is that my notion of an ideology
of emotion does not just include the ideology that an agent comes to hold due to
nurture in the form of acculturation and socialisation. It also allows for more
individualised versions, and for those individualised versions to be autonomous.
This may hold even where, as Wollheim (1999) argues of the moral sentiments, an
ideology was originally acquired heteronomously (see Chapter Two).

---

It should perhaps be clarified here that an idiolect is not something like a private language.
Rather, it is an idiosyncratic take on a shared language - the emotional equivalent of a distinctive way
of speaking or reasoning. Of course, an idiolect - of any of these kinds - may be more or less
impenetrable to others, which may raise significant meta-ethical problems when it occurs in moral
agency. I will return to this point, and its significance, in the next chapter.

This is not to imply, conversely, that agents who are not idiolectic are heteronomous. As my
argument in Chapter Two suggests, autonomy seems to be more a matter of how people came by their
attitudes and why they continue to hold them than the actual content of their attitudes. This "formal"
property of autonomy is also a main reason why autonomy, *qua* virtue, might be held to be merely
executive.

Also, while I shall not go into this point in any depth here, it is worth keeping in mind that
there may be group-based ideological subdivisions within any particular culture as well as between
cultures. Sometimes this is just a matter of brute diversity: Through contingent circumstances,
different groups have come to form different outlooks. Social and economic circumstances may make
a difference here, as may physical environment. Those from rural areas may for instance view things
differently from urbanites.
Thirdly, as indicated above, I will tend to include under ideological influences on emotion what might perhaps be thought of as more brute environmental or individual influences. The challenges posed by climate and other features of physical environment may for instance make a difference to ideologies. So may having lived through particular kinds of events, such as war or natural disaster. And so may more individual experiences or traits, such as undergoing significant illness, suffering handicaps, or having particular temperamental tendencies. But such in themselves brute features tend to become part of the ideologies held by the affected individuals. A tendency to become emotionally tougher, and perhaps more stoic, is for instance a common reaction to having undergone hardship. And this may be an ideological as well as a practical adjustment. So Shaw’s Alfred Doolittle, in Pygmalion, observing that lofty moral sentiments come at greater practical cost to the poor than to the rich, decides that he “can’t afford them, guv’nor”. But he also develops this idea into the more general claim that such sentiments are a luxury and an affectation, not just beyond his practical reach - a classic sour grapes move.

An important qualification: The impact of more or less contingent features of agents on ideologies should not be underestimated. But neither should they be treated in a deterministic manner. Given that humans are self-interpreters, and creatures of imagination, how and whether any particular fact features in their ideologies is not straightforwardly predictable: what turns one person stoic may turn another sentimental.

Fourthly, as I argued in Chapter Three and as suggested above, I take it that ideologies involve both descriptive and normative elements. These ideologies are not simply moralising accounts of what emotions we ought to have, relative to some...
normative standard. They are also accounts of what emotions we do have, and of what those emotions are like. And they may involve fairly sophisticated interpretations of the significance of our emotional natures being constituted in the particular ways that they are. But just as they are not purely moralising, neither are ideologies of emotion purely descriptive. And the normative concerns that feature in ideologies may bias as well as inform their descriptive elements.

I shall argue, below, that both the descriptive and the normative elements in ideologies may have a part to play in transforming our emotional natures. I will also argue that it may be of significance for how emotions is transformed, and how far such transformation is desirable, that normative and descriptive elements of ideologies of emotion may blur into each other. It is also, crucially, significant that both the normative and the descriptive elements may go wrong. And given that they inform and blur into each other, these elements may go wrong together as well as separately. First, though, some points about how ideologies feature in moral theories (and vice versa) may be in order.

**Moral theories and ideologies of emotionality**

Ideologies tell us what our emotions are like - but they also tell us what our emotions ought to be like. In the case of the lower-case stoic, the answer to the latter question is of course eliminated. Both ideologies of emotion and ideologies of emotionality are often significant elements in theories of moral agency. Not always: Some moral theories have relatively little direct interest in the emotions. Emotions often figure in utilitarian and deontological approaches only as brute phenomena to be taken into account, givens for moral agency to work around or in spite of. So they may figure in utility-calculations. Or they may be of concern for the deontologist insofar as they affect capacities for working out and carrying out moral duties. But they are seldom given a constructive role in moral agency - let alone an essential role. But even to take these kinds of views is of course to have some implicit theories about the nature and significance of emotions. And even moral theories that have little to say about emotion tend to influence ideologies of emotion, and of
emotionality - if only by cementing a tendency to marginalise emotions from moral and meta-ethical discourse.

Ideologies may however also play significant roles in meta-ethical approaches, and may even be significantly distinctive of the moral theories with which they are associated - if only in a negative manner. For instance, the Kantian account, as usually understood, has a negative general ideology of emotionality, viewing emotions as a class with a degree of suspicion, as being outside the bounds of rational control. For Kant, emotions tend to be what Griffiths (1997) calls irruptive: Disorderly, out-of-place upheavals in thought, motivation and behaviour; disturbances in what could otherwise be a stable and internally coherent rational order. Relatedly, emotions as a class are treated as irrelevant to moral assessment, and as not adding to the virtues of agents or actions. For having the right emotional inclinations, on this account, is a matter of natural virtue, which is a matter of luck in the constitution of one's empirical psyche. True virtue, by contrast, is a matter of reason and will, by which we transcend the contingencies of our empirical psyches. So the emotions are at best background noise, and at worst debilitating nuisances.

But the Kantian account, as Murdoch (1985) points out, also does seem to draw on at least some kinds of emotion, such as respect for persons and for the moral law, as crucial to good moral agency. (On Kant’s view of emotion, and comparisons with Aristotelian approaches, see also Sherman 1997, cited earlier). So we could see the Kantian account as having a positive ideology of those particular varieties of the emotion of respect. In consequence, Kantianism would seem to have a particular ideology of emotionality that sees some few chosen emotions as part of a rational bulwark against those emotions that will not follow the dictates of reason. We might of course still think that this exaggerates the badness of emotion, and that Kant’s insistence on control is informed by a misplaced belief that some ultimate rational control over self is possible or desirable.

We might also think, from a more Dionysian angle, that it is even desirable that emotions are not altogether subject to control or choice. Both in reflecting character and in reflecting reality, we assume that emotion represents something not altogether subject to will, and so may quite properly not be something over which
we have full choice. There seems to be something odd, morally as well as psychologically, about someone who claims their emotional reactions were radically freely chosen (see also Williams, 1973b). I have also argued, in Chapter Three, against radical voluntarism about emotion, both on descriptive and normative grounds. This is not, of course, to deny that some degree of control over emotion is both possible and desirable. And even within a broadly Apollinian framework, we might be able to give an account of how other emotions, too, could be brought within the moral realm, and not just be seen as messy, out of control, irruptive states.

Some writers, like Aristotle, extend this possibility to emotions generally, and suggest no emotions are in themselves morally bad or irrelevant to moral assessment of those who feel them. Others (Hume, Smith) operate with a narrower range of specifically moral emotions such as compassion, sympathy, and benevolent feelings. These emotions are triggered by and constitutive of moral concerns, where other emotions such as fear or anger may be triggered by concerns that are not specifically moral. The moral emotions are also typically “corrected sentiments”: Getting them right involves practice, experience and discernment. And we could perhaps class with these narrower conceptions of moral emotions the Kantian account as sketched above, and broadly Kantian approaches to moral sentiments such as that of Rawls (1971: especially Sections 73-75). Of course, the range of what is thought to be moral concerns is itself a contentious issue, independently of emotion-related issues. And different moral communities may, quite apart from any kind of philosophical affiliations, vary in which among the designated moral emotions they focus more strongly, and how these emotions are thought to relate to

---

7 Aristotle does seem to think that some emotions, notably envy and malice, are in all cases bad. But he suggests that these may still be seen as perversions of emotions he considers more morally acceptable. So envy might be seen as a perversions of the urge to emulate what one finds admirable, and of righteous indignation at goods being enjoyed by the undeserving at the expense of the deserving. And malice might be seen as a perversions of the (to Aristotle) acceptable satisfaction we feel at the wicked getting their comeuppance. So when applied to the right objects and for the right reasons, even otherwise bad emotion may on this account not be a problem. See for instance the discussions in NE 11.5, and Rhetoric 11.9-11.

8 Hume’s notion of the general point of view as a corrective to the distortions of particular perspective, and Smith’s notion of the impartial but benevolent observer as the ideal moral judge, both point to this “corrected” nature of moral sentiments. Adopting these points of view acts as a corrective to the possible biases - and honest mistakes - that sentiments of natural virtue are liable to.
each other⁹.

On the whole, as will have been apparent, I am inclined more towards the Aristotelian view. I have argued (Chapter One) for the specifically moral emotions to be a subset of the participant reactive emotions, which in turn are a subset of emotions generally. And I have argued that there are crucial conceptual connections between these emotions and some of our most basic intuitions about moral status, moral awareness and moral motivation. In all of these, I have argued that the specifically moral participant reactive attitudes have a crucial part to play.

But I have also argued that emotions that are not specifically, inherently, moral may have a crucial role to play in informing moral agency and in informing more specifically moral emotions. Insofar as our non-moral emotions reflect what we are like and under what conditions we can flourish, they may also legitimately constrain what kinds of aims morality ought to promote. And the specific way an agent feels non-moral emotions, such as anger, reflects significantly on their moral character. What makes someone angry, and how and whether they express and act on their anger, will tell us a great deal about them as a person and as a moral agent. So will any meta-level reactions they have towards their own anger.

But conversely, the non-moral emotions themselves will on this view be constrained by moral considerations: We may want to put limits on what kinds of anger, fear and so on can be morally permitted. And the more specifically moral emotions may play such a constraining role relative to the non-moral emotions. Where emotions are viewed as crucial to moral agency, the role of the specifically moral emotions may be viewed as one of ruling both other emotions and other faculties such as reason or will - in Baier’s (1980) terms, some kinds of emotion are moral “master passions”. We might for instance think that anger is only morally permissible insofar as it is compatible with an overall attitude of compassion, benevolence, justice or respect for persons. But within such constraints, non-moral emotions like anger may play an important informative role in alerting us to transgressions. That we feel angry at how someone treats people may alert us to their

⁹ See e.g. Rozin et al. (1999) on cultural variation in the relative prominence of anger, contempt and disgust as moral emotions and how these variations correlate with differences in moral outlook. On social sources and influences on particular emotion-types, see also Rorty (1998).
behaviour being morally wrong.

A significant part of how ideologies figure in theories of moral agency is through assumptions about how we come by emotional skills. For a Kantian, the inclinations appear much too vulnerable to contingent features of the context of learning to be viable virtue-bearers. Aristotle's own works may be thought to concede this point, as he explicitly does not think emotional virtue is possible for all, and emphasises the crucial importance of formative influences in shaping emotional dispositions. A very basic challenge for modern writers of an Aristotelian inclination is how to reconcile a central role for the emotions with concerns about their vulnerability to contingent formative influences. One solution might be to co-opt some degree of Kantianism and insist on the capacity for agents to transcend the contingencies of their original context of moral learning. We might also want to adopt a more optimistic approach to the possibility of life-long moral learning, de-emphasising the importance of early habituation. And of course, most modern Aristotelians also take a more “equal-opportunity” approach to virtue than Aristotle himself, discarding the idea that gender, ethnicity, social class, and climactic conditions are relevant to capacities for developing moral virtue. I will not argue the factual case (such as it might be) for these adaptations of how moral agency is viewed in depth here. But I do think that, taken together with a less brute view of emotion, these adjustments could allow for a viable compromise between Aristotelianism and Kantianism about the role of emotion in moral agency.

That emotions are amenable to ideological adjustment may then give hope that we can bring emotions as a class within a moral framework. And the possibility

---

10 Though see Fortenbaugh (1975) who while he resists the view that non-Greek men are less rational than Greek men, is rather cavalier about worries that Aristotle represents women unfairly. While these points are not central to Fortenbaugh's otherwise very useful analysis of Aristotle, they are almost the more worrying for commanding so little reflection on Fortenbaugh's part.

11 The Politics, VII.7, explicitly states that the Greek climate is ideal for producing high-spirited and intelligent men. Northern Europe, by contrast, produces people of high spirit but lacking in intelligence, while Southern Asians have intelligence but lack spirit. By some amazing luck, Aristotle finds himself once more in the better category of agent.

12 Interestingly, other emotions than the specifically moral ones might also be viewed as ones that have increasingly become subject to “equal opportunity”. For instance, romantic love, boredom and melancholy have all historically tended to be “elite” emotions, indulged by a leisured and usually educated minority. (On the history of boredom and melancholy, see e.g. Svendsen, 1999). These days, at least in Western culture, these emotions are ones that anyone might be expected to have experienced.
of overall ideologies of emotionality also suggests some grounds for hope that we can do better than piecemeal ideological adjustment of particular emotion-types. Some ultimate moral theory would perhaps also provide a coherent and realisable standard for what our emotional natures should be like. Assuming, of course, that such a moral theory is itself a coherent possibility.

But it also seems reasonable to assume that human nature and the human situation impose some constraints on how far ideological adjustment can successfully go. As I will argue below, there may also be significant internal connections between different emotion-types, and these connections increase the challenge to any attempt to reshape emotionality. Moreover, the particular natures of individual agents and their situation impose constraints on any universalising ambitions to make our individual emotional natures more alike - a point I will follow up further in the next chapter. And as already suggested, constraints of these various types may not even be a bad thing. For if morality is significantly about furthering human flourishing, attention to actual features of human nature and the human situation can hardly be dispensed with. Which is not to say that determining what those features are will be straightforward. Nor is distinguishing what is a product of ideology and what is “natural” straightforward - it may not even be a coherent project. And I will not try to do full-scale disentangling here. But some crucial points about the way emotion is shaped by ideology can still be made, short of full disentanglement.

4) How ideologies may transform emotions - and some caveats

Elements of ideologies

To recap some of the elements of ideologies before considering how they could and should influence our actual emotions: In Chapter Three, I included in the ideology of an emotion expectations about the typical formal features of the subjects, objects and grounds of the relevant emotion-type. So ideologies tell us

---

13 Given that the cognitive paradigms for emotions may be merely quasi-intentional, “formal” features may be understood in a relatively brute way. For instance, an animal’s reacting with
what people typically think, want, and feel, and why, when they have particular emotional reactions. They tell us what kind of objects particular emotions are likely to be felt towards and about, and what it is about those objects that make this so. And they tell us what the emotions tell us about those who feel them, and what kind of emotional reactions we can expect from which kinds of persons. And they will probably tell us something about the relation between occurrent felt emotion and emotional dispositions. Emotional dispositions, in turn, have important relations to the character and temperament of agents.

I have argued that these expectations can be captured by loose conceptual paradigms. These ideological paradigms aim to represent the typical cognitive and attitudinal or motivational content of the emotion itself. In the paradigm, we could also include expectations about how the emotion is expressed in subsequent attitudes and behaviours. Behaviours may include the merely expressive ways in which the emotion is manifested in facial and other physical displays - trembling with indignation, beaming with happiness, frowning in puzzlement, and so on. In other words, the paradigm tells us what kind of inputs can trigger the emotion, and what kind of outputs may result from its being triggered.

The ideology of an emotion also indicates how an emotion may be identified as being of a particular type, even where it is on the margins of the paradigm. We may know that we are angry, for instance, even where the anger manifests itself in atypical behaviours - making the person burst into laughter, for instance - or is directed at a surprising target. But it is also significant, as I shall argue at greater length below, that ideologies may draw the paradigms of emotions so as to make us misinterpret or overlook such cases.

The normative elements in ideologies of emotionality range from relatively thin normative standards, such as minimal rationality or intelligibility, to the very thick such as moral rightness or wrongness. Somewhere in between on the scale of normative thickness are standards of rationality and appropriateness. How a particular kind of normative standard features in a particular ideology, and how prominently, varies considerably - largely, I would suggest, depending on how aggression, concupiscence or avoidance towards objects of a particular colour or shape may be a
responsive emotion is thought to be to such normative concerns. If emotion is thought to be random, unpredictable, unintelligible, fundamentally non-rational, there would be little or no recognition that emotion may be measured by and respond to normative standards. Also, as I have suggested earlier, the line between normative and descriptive elements in ideologies of emotion may be significantly blurred, in both directions. The view that emotion is non-rational may for instance feature both as a descriptive claim and as a normative judgement on emotion.

Ideologies of particular emotion-types may also be influenced by both general and particular ideologies of emotionality. This is perhaps especially true of normative issues. It may for instance be part of a particular ideology of emotionality that one should react to transgressors with forgiveness or sympathy rather than anger and vindictiveness.14 And of course, the ideology applied to a particular emotion-type may in its turn inform the associated particular ideology of emotionality. If we think that anger is a bad emotion, that will obviously crucially inform how it figures in our particular ideology of emotionality. Whether our starting-point is emotions as a class or particular emotions may in turn depend on the associated general ideology of emotion. If emotions are viewed as a peripheral part of our lives and psychologies, then we may not bother to make explicit finer distinctions between them. In this case, a particular ideology of emotionality may at most be something we can infer from collating ideologies of particular emotion-types. On the other hand, if emotions are viewed as crucial, we may be particularly concerned with how different emotions relate to each other, and how each emotion-type plays a part in informing agency.

Ideologies of emotionality, then, provide a crucial background meta-level to the ideologies of particular emotions. Assumptions about emotion such as those that inform the objections to emotion (cf. Chapter Two) may figure in such meta-ideologies. As, of course, will the kind of assumptions I have drawn on in building a meta-ethical defence of emotion. So in addition to telling us about the typical objects, subjects, grounds and so on involved in an emotion-type, ideology also

---

14 While this might be thought a distinctly Christian view, it seems perfectly possible that adherents of other religions, as well as atheists and agnostics, could take this view.
involves meta-level assumptions about how and whether emotion is controllable, rational and so on.

**Ideologies and "natural" emotion**

There is considerable variation in all these elements of ideologies, both between cultures and between individuals. And while we might have doubts that the variation in reality is as great as the variation in ideology suggests, there is both scientific and anecdotal evidence that there is variety in how emotions are felt and understood. Moreover, ideologies of emotion penetrate deeply into even those emotions that are the most obvious candidates for being nature-given, such as fear and anger. What we are angry about, or afraid of, and how we express those feelings, is open to considerable ideological influence. And ideologies may cause variety in emotional reactions, as well as reflecting pre-existing variety: Both the descriptive and the normative assumptions that people make about emotions may affect the emotions they actually have.

And ideologies may not just shape our emotions: some emotion-types, such as patriotism and romantic love, have been argued to owe their very existence to ideology. By this I mean that ideology does not just play a part in directing the emotion towards particular kinds of objects, or in modifying the way in which it is expressed, but that the emotion might be thought to have no ideology-independent existence. It has been argued that people would not fall in love if the idea had not been suggested to them: contentious though that claim may be, it seems more intuitively plausible than a similar claim made about anger or fear. Some have thought that our more specifically moral emotions, such as a sense of justice or a

---

15 For more in-depth discussion of "nature vs. construct" issues about emotion, see e.g. Griffiths (1997) and more or less any of the essays in Harre and Parrot (1996). Related issues are also discussed in e.g. Elster 1999a, in Ekman (1977, 1994) and in Ekman and Friesen (1989). For criticisms of overstated versions of the natural/constructed distinction, see again Griffiths, op.cit., and Goldie (2000a, Chapter Five).

16 Which is not to deny that both of these emotion-types have particular kinds of instances that seem unlikely to occur without ideology to encourage them. See for instance Griffiths' (1997) discussion of culturally sanctioned ways in which someone might be recognised as in the grip of uncontrollable anger, such as "running amok" in certain Asian cultures or what he calls "Rambo syndrome" in Western culture. These forms of emotional display are notable for diminishing in
respect for persons, are such ideological artefacts: not natural to the individual, but instilled in them through socialisation. This might be one way of reading Wollheim’s (1999) claims about the heteronomy of moral emotion. Though of course this is itself a contentious issue. As is the question of what would follow from moral emotions not being natural. And the notion of nature is of course itself problematic and multiply ambiguous: do we, for instance, include in our nature only what is given by our genetic coding, and view anything that results from the impact of the environment, learning from experience included, as in some sense artificial?

It should be noted, however, that even if we view some of our emotions as being ideological artefacts in this way, it need not straightforwardly follow that they are not important or genuine. The notion of genuineness is of course itself significantly vague. For humans are creatively self-determining creatures. We do not just choose between the options laid out for us by our nature and our environment, we add options of our own making. And even those capacities that are in us as givens may owe significant part of the particular form they take to learning through socialisation and more brute environmental influences.

So even if moral capacities are ones we have to learn, or they otherwise involve modifications of our innate natures, whatever those might be, it does not follow that they are not part of our human nature. For human nature is in large part second nature - a matter of learning and habituation, in which creative self-determination also plays a significant part.

Some ethicists, like Aristotle, explicitly embrace a view of moral virtue (and by implication vice) as being a matter of second nature. “Neither by nature, then, nor contrary to nature do the virtues arise in us; rather we are adapted by nature to receive them, and are made perfect by habit” (NE 1103a23). Aristotle also argues, just before this, that “it is also plain that none of the moral virtues arises in us by nature; for nothing that exists by nature can form a habit contrary to its nature”.

As applied to emotions, that last point would suggest that emotion as given by nature may be capable of relatively little adjustment - that it cannot take a form "contrary to its nature". And there appears to be evidence that at least some of our frequency when social and legal toleration of them is withdrawn.
emotion-types are ones we are biologically hard-wired to have. These are often referred to as basic or primary emotions. The basic/primary emotion-term in itself is somewhat misleading, since it is also applied by some to emotions that are counted as primary on the basis of analytical or conceptual reduction rather than empirical evidence from biopsychology or anthropology. And within either of those two approaches, there are significant differences as to what emotion-types are counted as primary, and about the exact grounds. As Elster (1999a: 243) notes, while certain emotion-types tend to recur across different lists of basic emotions, some authors do not list fear and anger, both of which are held by most writers to be paradigm examples of basic emotion-types. Love, in the general bonding- or attachment sense rather than the romantic sense, appears in some lists but not others.

More immediately significant for my purposes than the question of which exact specific emotion-types can be counted as basic is that even those emotions that are typically listed as such display considerable diversity and plasticity when it comes to specific manifestations. For instance, the capacity for feeling fear might be hard-wired into us, and present in all cultures. And some of its outputs, such as the associated facial displays and patterns of neurophysiological arousal, are not subject to voluntary control in their most immediate stages. They might also be fairly stereotyped: Ekman argues that facial displays of fear, as well as other basic emotions, are strikingly similar across cultures. But what we are afraid of, how we deal with fear, whether we aim to suppress or control it, and how we think fear reflects on us, may be significantly subject to cultural and individual variation. So even if there are some emotion-types we have no real choice about having dispositions towards, it does not follow that we are hard-wired to react to particular objects in fixed ways. So nature may be more ideologically permeable than expected. In fact, ideology and its various constituent elements may be what gives emotions, or at least human emotions, particular shape at all - nature may just provide potentialities and set the perimeters that ideology operates within,

17 For examples of the former (what Strawson (1979) calls conceptual-analytical accounts), see e.g. Descartes (1984), Spinoza (1677). Spinoza's account is particularly striking in that he reduces emotion's basic components to pleasure/displeasure and desire. For examples of the latter (genetic-psychological accounts), see e.g. Ekman, as cited above, Arnold (1960a, 1960b), Frijda (1986), Damasio (1994, 2000) and Griffiths' (1997) theory of affect programs.
significant as both of those are.

Of course, not all writers on emotion believe in basic emotions - Lazarus and de Sousa, for instance, both query the existence and significance of such emotions, on scientific and philosophical grounds respectively. We might think that basic emotionality would just come down to our possessing some more or less plastic capacity for feeling, where this capacity would manifest itself in different ways depending on what we are like and what our situation is like, and to some degree on choice. My argument in the previous chapter could be seen as an outline of such an account.

Now even here, certain basic constitutional features of humanity may shape and limit what our emotions are capable of. And these features also constrain such transcendence as we are capable of, emotionally and otherwise. There may also be limits on what shape emotion, and emotionality, can take in terms of the internal logic of affect. There may be significant internal connections between the cognitive contents of different emotion-types that limit our capacity for picking and choosing between them. Some emotion-types might be particularly rich in such entailments. Attachment to others, for instance, brings with it a particularly broad range of dispositions to other emotion-types - concern and fear for their welfare, vicarious participant reactive attitudes towards third parties, jealousy of their affections, and so forth. But even if other emotion-types may not be as rich in entailments, virtually none of them seem likely to stand altogether alone. Dispositions to anger or fear, for instance, seem unlikely if we did not have affect-bound concerns that could be threatened.

We might also think of this "logic of affect" in terms of dramatic or narrative coherence in the cognitive content of an agent's emotions - among others, de Sousa, Wollheim, Nussbaum, Hepburn, Goldie and Elster have all made suggestions on these lines. Of course, human temperament is capable of considerable variation: I do not want to suggest that there is some uniform logic of affect operating in all agents. But I do want to suggest that our emotional natures, both on a species-level and on an individual level, may exhibit an internal, *sui generis* logic that significantly limits our choices. Within the constraints of this logic, however, we have considerable
room for manoeuvre. (That there is such an internal logic to emotion is of course itself an ideological assumption, but hopefully a warranted rather than distorting one.)

5) Ideological shaping of emotion - and some caveats

Ideology as (mis)representation

In their descriptive aspect, the ideologies inform - in fact, they are - our understanding of emotion and what it does to us. I have suggested, above, that ideologies serve to fill in the particulars of what our emotions are like, subject to the more or less plastic constraints set by human nature, the human situation, and the logic of affect itself. In their normative aspect, they express approval or disapproval of our emotions, relative to some standard of ours.

Some very basic ways in which ideologies of emotion may change emotion follow fairly straightforwardly from these points. Emotional learning - our coming to understand our emotions - takes place in a context set by ideologies of emotion. Even the most seemingly automated and instinctive reaction in an adult agent will probably be significantly informed by the ideological context in which that agent's emotional learning took place. And this ideological context is not an innocent representation: it need not straightforwardly capture what our emotions are actually like. It need not even straightforwardly capture what ideology makes them be like. It may for instance be a significant part of an ideology of emotion that the way our emotions are shaped is given by nature, and not subject to our choice.

And at the same time, since ideology provides the means to emotional understanding, it may have a powerful self-fulfilling effect. So for instance, someone who has been raised to believe that emotions are uncontrollable will be less motivated to try and control their emotions, simply because they do not believe such attempts are likely to succeed. Such mistakes may be more or less motivated. Sometimes, they may just be honest mistakes. But often, particularly since ideologies have a strong normative component, they may be motivated mistakes -
we may represent our own emotions, to ourselves and others, as more reasonable and morally correct than they really are.

It may often also be that the mistake is not so much due to the agent themselves getting things wrong, as to a socially held ideology - we could call this category heteronomous mistakes. There are considerable cultural differences in how and whether particular emotion-types are recognised in language and in emotion-talk (Heelas 1996). Some cultures may not lexicalise (name) particular emotions, though that need not mean that agents in that culture are unfamiliar with the emotion itself. For instance, someone need not be a German-speaker to experience Schadenfreude. As Heelas also notes, there are considerable cultural differences in how emotion-types are cognised - how prominently they feature in the emotion-talk of the culture, or as we might say here, in the emotion-ideology of the culture. Romantic love, for instance, is hyper-cognised in our culture. By contrast, the emotion the Japanese call amae - a kind of happy sense of dependency - is hypo-cognised, and not lexicalised, in current Western culture, though we might not think that Westerners are necessarily strangers to the emotion itself.

Social ideologies may then more or less innocently misrepresent emotional realities. But they may also distort them in more motivated ways. For instance, most cultures make more or less overt assumptions about differences in emotional virtue based on gender and ethnic background. As indicated above, Aristotle makes assumptions about differences in emotional virtue based on gender, class, ethnicity and cultural background that strike most modern readers as rather astonishing. Critical reflection, however, tends to suggest that such views involve motivated distortion. For one, the group making the assumptions almost invariably comes out best. And these distortions are not harmless - they inform the way in which people are socialised, and may leave them with fundamentally wrong descriptive and normative assumptions about themselves as well as about others.  

---

18 Of course, alternative ideologies proposed to correct the distortions may themselves be descriptively and normatively questionable. Lutz (1996) for instance queries the descriptive accuracy of "pro-women" assumptions about gender differences in emotionality such as those proposed by Gilligan (1982), and a number of writers, feminist and otherwise, have taken issue with Gilligan's take on particularist normative assumptions. (See for instance the discussion of Gilligan in Flanagan, 1991).
Some points more specific to normative standards

That we measure our emotions by normative standards also makes a difference in less underhanded ways. Not necessarily directly to what our emotions actually are. We may disapprove our emotions and still have them, as anyone attempting to overcome phobias or prejudices will know to their cost. And we may signally fail to have emotions that we wish we had, as anyone who has suffered a loss of motivation knows. Of course, we might expect that a degree of heteronomy in formative stages of our emotional learning may tend towards making us have the “right” emotions by the standards of those who taught us. In the course of growing up, for instance, we learn to find disgusting, guilt-inducing and shame-making things that small children are quite unfazed by, such as the urge to clobber people who get between us and what we want. But we may find ourselves adult moral agents and still have feelings that neither our society nor we ourselves approve. We may also, by this stage, have developed emotional ideologies of our own that diverge significantly from those of our society. Our actual emotions may be in line with our own standards, the society’s, or in particularly unfortunate cases, neither.

But while our normative standards may fail to wholly determine what our emotions actually are, whether we approve our emotions or not makes a difference to whether we try to control them or change them by such means as we have at our disposal. And it also makes a difference to whether we get behind them. To put it rather paradoxically, we do not always identify with ourselves emotionally. Our feelings may go against what we believe, against what we wish we felt, and against what we would want to believe ourselves like - I have discussed this earlier under the heading of internal heteronomy. In such cases, we may try to dissociate ourselves from the emotion, or from some part of ourselves presumed responsible for the unfitting emotion, in favour of sustaining what we believe or wish to believe about ourselves.

Or we may try to change what we feel. Where appeals to what it would be proper to feel in terms of the formal object and grounds of an emotion fail to rid us of the embarrassing feeling, we may try to use our understanding of the formal subject of the emotion. This may involve moves similar to those we make in trying
to understand or share in the emotions of others (see Chapter Four), but in this case
with our ideal self as the one we aim for emotional identification with. So virtue
theorists of an Aristotelian bent might advise us to identify with the virtuous person
we could be in order to feel like, act like and ultimately become that person.

As a means of changing character, this approach crucially draws on
ideological assumptions about how character and emotion are interrelated. This
strategy may of course be limited by how far our actual personality is from our
imagined ideal. There may be some gaps that cannot be bridged psychologically - or
at least not without great effort. And there may be limits to how far we are willing or
able to make the necessary changes in practice. For instance, the emotional life of
Thurber’s Walter Mitty is that of his imagined heroic alter ego, but for all his
emotional identification with this alter ego, he remains just Walter Mitty in his real
life and actions.

Such moves may also be made in a less global way, to change some
particular emotion rather than our entire character. This approach is usually known
as bootstrapping. Bootstrapping may be more or less conscious. Unknowing
bootstrapping may for instance occur where a person initially misidentifies their own
emotional state, for instance mistaking a more or less accidental mood for a feeling
about a particular object or person. Acting on what they believe to be their emotional
state, they may then induce the emotional state they wrongly believed themselves to
be in. Knowing bootstrapping occurs when the person consciously induces the
desired emotional state by feigning it.

Bootstrapping, along with other methods of controlling and changing
emotions, may significantly increase the extent to which our emotional life is under
our control. But we would expect that there are limits to how much a bootstrap can
hoist, and how far, even with practice on the part of the bootstrapper. And while we
might applaud the outcome of bootstrapping, we may also have legitimate
misgivings about both the ends and the means. We might worry about the
desirability of the outcomes of bootstrapping - as indicated above, the normative
contents of ideologies of emotion may be suspect. And if they are suspect, so would
the outcomes of adjustments informed by them be. Part of the misgiving could
however also come from the role a degree of passivity in our emotions has in notions of sincerity: I have discussed earlier (Chapters Two and Three) the problems raised by voluntarism both regarding emotions and regarding valuations. Bootstrapping, and other forms of emotional adjustment, may obscure to us what our real feelings are, which in turn may keep us from crucial information about both ourselves and the objects of our emotions.\textsuperscript{19}

6) The pros and cons of ideological adjustment

To sum up, then: This ideological dimension may then be something of a two-edged sword. Our ideologies of emotion and of emotionality reflect our capacities for interpreting and determining ourselves - which are also crucial moral skills. But the ideological dimension also reflects our capacity for deceiving and distorting both others and ourselves.

Worries about distortion apply most obviously at the level of representation: Ideologies of emotion may misrepresent the realities of emotion. Through misleading ideologies of emotion, we may misidentify or fail to identify feelings that we actually have. But since ideologies also influence the realities that we, as self-determining and world-changing creatures, bring about, we might have worries about distortion that go beyond mere misrepresentation. Particularly so since ideologies tell us what emotions we ought to have. Ideology may play a part not just in making us think that we ought to have emotions that are actually morally bad (xenophobia, for instance), but in making us actually have those bad emotions. And of course, the distinction between the descriptive and the normative may be blurred here. Wanting to think ourselves good, we may engage in motivated self-deception about what our emotions really are, misrepresenting them as closer to our normative ideal than they actually are, and possibly damaging both ourselves and others in the process.

So the opportunities that our capacity for ideological adjustment of emotion

\textsuperscript{19} For more detailed discussions of possible "pathologies of bootstrapping", see e.g. Midgley (1981), Elster (1983, 1999a); de Sousa, (1987, Chapter Nine); Pugmire (1994, 1998 Chapter Eight).
brings come at considerable potential costs. And the costs are side effects of the same capacities that provide the benefits. I am inclined to think that the benefits outweigh the costs, at least potentially: if we did not have these capacities, we would be capable of very limited if any moral agency. But I would not like to say categorically that the actual benefits outweigh the actual costs. At any rate, it seems useful to have some idea of the potential pitfalls of ideological adjustment of emotion, and of the limits on its usefulness.

The notion of ideological adjustment as a possible source of distortion also echoes the point, raised earlier, that reason as well as emotion may be a source of error. If the ideologies that reshape our emotions are informed by reason, and reason may go wrong, then our emotions will by extension be reshaped in distorted ways.

Ideology may also cast emotional reactions that do not conform to the ideological standards as wrong. In these cases, ideology has not fully penetrated the emotion, but has penetrated the person’s attitudes to their own emotions. And while a capacity for such critical reflection is essential for moral performance, it may not always be critical reflection that gets it right. Sometimes, a morally good instinct may get sicklied over with the pale cast of thought. And the worry is not just that reason could reduce us to Hamlet-like inaction when action is crucial. As Bernard Williams has suggested, one thought too many may lead to wrong action as well as to absence of action. This is not to say we should give up on critical reflection: On the contrary. Nor should we just let instinct rule. But neither should we get too optimistic on reason's behalf when it comes to getting moral agency right.

The issues in this chapter also tie in with the suggestion, raised earlier, that what is at stake about emotion in moral agency may not be specific to emotion as such. Rather, emotion might be used as a McGuffin for underlying tensions in our meta-ethical thinking. Specifically, there are worries about how what is given in our nature and situation should feature in our moral thinking. And conversely, there are worries about what kinds of modifications of our natures would be possible, and

---

20 See again Bennett (1974): for a slightly different take on a case where moral zeal subsequently caused the overreaching zealot intense emotional distress, see Langton (1994) on Kant's correspondent Maria von Herbert.
what kinds of modifications would be desirable from the point of view of morality\textsuperscript{21}. That our capacity for adjusting emotions may not just be a good thing also suggests that worries about emotion may be traced to general worries about our abuse of freedom and reason, as well as to worries about lack of control and reason. So while traditional objections to emotion often emphasise emotion as being non-rational and out of our control, the very cases in which emotion is led by reason and self-control may raise concerns about emotion. I have also argued that the critical perspectives that we adopt on emotion may themselves be significantly misguided, both descriptively and normatively - as I argued of the lower-case stoic’s position in Chapter One.

Here, I have been concerned with these issues in their more general form: How nature might determine emotion, how ideology might adapt it, and how this ideological dimension relates to morality. In the next and final chapter, I will focus on some issues our individual particularities might raise for morality, and how emotion figures in this.

\textsuperscript{21} I will have relatively little to say about the notion that there might be desirable adjustments of human nature that would be impossible. Granted, individual psychopathology may be so extreme as to raise real worries about whether the person can count as a moral agent at all. But I do not think it is plausible to extend this worry to humanity as a species, imperfect though we demonstrably are. And as my argument here and earlier suggests, I am sceptical about the desirability as well as the practicability of radical adjustment of human nature.
Chapter 6: Emotion, difference and moral agency

1) Introduction

In this final chapter, I examine some worries about emotion in moral agency arising from the way differences in emotional reactions and dispositions reflect and even cause differences between agents. I want to argue, again, that these problems arise largely independently of emotion. Rather, as I suggested in Chapter Four, the problem lies in the nature and situation of agents. More specifically, what is at issue is the diversity of agents, and how such diversity bears on capacities for moral agency. Emotion, and human nature, could theoretically be more uniform: It is the fact that human nature takes diverse forms across agents that raises meta-ethical worries. Insofar as there is a problem, that is: I also want to argue that within certain limits, diversity is permissible and even desirable.

I will begin by setting out how the issue of diversity bears on the kind of worries that have been raised here about emotion in moral agency. In the light of this, I then look at a worry for diversity: The possibility that morality imposes uniformity across agents. I point to some pressures towards uniformity: Demands for universalisability, the notion of morality as rule-governed, and, perhaps surprisingly, some ways of reading the doctrine of the unity of the virtues. I argue that there are general reasons for resisting pressures towards uniformity out of concern for human flourishing. But there are also reasons to query the specific types of pressure that incline towards uniformity. The end result is a dilemma. Specifically, I argue that the degree of internal coherence which is not only necessary for an agent to flourish, but also for him to perform well as a moral agent, may be crucially tied up with his distinctive traits. But the pressures towards uniformity make distinctive traits problematic. While this might seem to be a problem about emotion (through its association with particular agent-relativity), though, it in fact turns out to be a problem caused by antinomies in our assumptions about moral agency. In the light of this, I conclude again on the McGuffin theme.
2) Diversity and the objections to emotion

Subjectivity and the subject/object-balance in emotions

I have argued, in Chapters Three and Four, that emotions reflect not only their objects, but also their subjects. Emotions do not just reflect what those things (or persons) are like that the emotions are about or directed towards. Emotions also reflect what those that have the emotions are like. And what those that have emotions are like is not a simple question: human nature, and the situation of particular human agents, often takes diverse, complex, and elusive forms.

I have argued earlier (Chapter Two) that subjectivity in what I have called the phenomenological sense need not raise meta-ethical problems: subjective experience can theoretically be seen as merely a correlate of objective reality. Nor, I argued, need projection necessarily be bad subjectivity. However, there are some general points about subjectivity in these senses that will be worth raising here for the significance they gain when associated with the kind of particular agent-relative subjectivity that is at stake in relation to diversity. So I shall start here with that more general issue.

As I suggested at the end of Chapter Three, when introducing the notion that subjects as well as objects shape emotions, the balance between the importance of the subject and of the object in emotion may be a difficult point to work out, both descriptively and normatively. In the case of any particular emotional reaction or disposition, it may be difficult to say whether it reflects the object or the subject more, or indeed whether it reflects the two equally. Whether emotion in general tends more towards reflecting subjects or objects is an even more elusive point. Nor does it seem clear what the balance between reflecting the subject and reflecting the object ought to be. What we love, hate, fear, long for, tells us much about both ourselves and about the objects of our emotions. But it is not immediately obvious which of the two they tell us more about, or which they should tell us more about. My main concern here is with the normative issues, but the descriptive issues will obviously also have some bearing on how that turns out.
The balance between subject and object will not always be a matter of great or obvious concern. But in some cases, notably intersubjective ones, the balance may be a matter of acute concern. Where the object of emotion is another person, with their own point of view on the world, on us and on being the object of our emotion, we would think emotions towards them should take all those facts into account. How this should be done, and to what extent, is of course still a matter of debate, and intuitions may vary from case to case. It may for instance depend on the moral standing of the other person themselves, or on the type of emotion felt. We may feel justified in taking little account of the point of view of those we despise, but hardly of those we love. But for someone to be the object of either of those emotions in the first place will of course typically (and properly) entail that we have some opinions about the moral or other worth of that person in themselves.

**Subjective and objective as a spectrum**

To see more clearly what is at stake here, we can imagine this as a matter of locating emotion on a spectrum. On one end, we have emotions that reflect only their subjects: total projection. Such emotions might entirely lack specifiable objects, or objects might feature only as more or less randomly chosen targets for emotions which they have no real causal role in arousing. Such emotions would be purely subjective in the sense of being only about the subject of the emotions. They could of course still present as being about something other than the subject who feels them, but this would be a false appearance. The subject themselves may be more or less clear about their emotion being projective, and may view such projection with a greater or lesser degree of approval. Some might embrace projection as a particularly sublime expression of (their own) particular subjectivity - some versions of Romantic celebration of sentiment being possible cases in point.

At the other end of the spectrum would be emotions that reflected only their objects: emotion as pure observation, without any shadow of observer effects. Such emotions would be in a sense purely objective, in the sense of reflecting only the object that they are about. In such emotion, no personal or partisan interest would be brought to bear, nor would there be any trace of the perspectival. This is fairly
obviously only a theoretical possibility. Though perhaps some particularly godlike creature (assuming its existence is a coherent or real possibility) could pull off this emotional feat. Or perhaps such emotion could be felt from Nagel’s view from nowhere.

Barring some particularly radical form of solipsism on the part of the subject of the emotion, purely projective emotion may also be only a theoretical possibility. And possibly not a very coherent notion, at that. For it seems an implicit requirement that there be something, distinct from the observing mind, for emotion to be directed at, and that the emotion reflect salient features of this object as well as of the subject.¹ Still, these two rather implausible extremes of total objectivity and total subjectivity will serve to set the poles of the spectrum. In between are emotions that reflect both their subjects and their objects, but to varying relative degrees.

Now, assuming that pure projection and pure observation are both unrealisable options, the question of whether either is desirable becomes a moot point. Though the answer, at least in the case of projection, seems to be a fairly obvious no. So where should we aim to find ourselves on the scale in between?

There does not seem to be an obvious response to this - or at least not one that would not be in need of some qualification. As I have argued earlier (Chapter Two), subjectivity as such may not be a real problem, and may in fact be unavoidable. We might also think that moral issues - and perhaps moral values - only exist insofar as there are creatures that they could be moral issues and values for: in this sense, morality may be an inherently subjective matter. Though as I have said earlier, I will not argue this particular point further here.

But while the existence of observers in some abstract sense might be permitted and even required for moral issues to arise, admitting particular observers whose particularity would shape the content of their observation raise further contentious issues. For if agents are many and different, it seems the same object may be different depending who perceives it. And that seems a problematic outcome - at least on some readings of the requirements for moral assessments to be objective

¹ In the extreme solipsist case, where the observing, emoting mind is both the subject and the only possible object of emotion, we can still make at least a conceptual distinction between the subject and object, and hence identify an object.
and rational. I shall look in more detail below at how some common intuitions about the nature of morality put the case for diversity under pressure - and how some problem cases may help towards making a case for diversity.

**Emotion, difference and internal coherence in agency**

On a typical universalist approach, particular features of the subject making a moral judgement should not colour their judgement. If observers are particular in their nature, possessed of distinctive, idiosyncratic traits, located in particular positions within particular situations, their observations should as far as possible be cleansed of all those forms of particularity. So one could make valid judgements from the general point of view, from Rawls' *original position*, or after applying some test like the categorical imperative. A valid judgement, on this reading, is one that any rational person could share, regardless of their particular features or situation, and regardless of their position within a situation. From a relativist point of view, by contrast, the morally correct response may depend to a greater degree on specific features of the subject. For a cultural relativist, the features in question would be generic; for a more individualist form of relativism, it might be specific to the individual in question. For a particularist, depending again on whether they tended towards individualism or not, it might be either.  

Of course, some kinds of emotional reactions are in themselves often fairly generic, such as those involved in reacting to perceived threat to one's loved ones. (I am assuming, at this point, that particular attachments are morally permissible in the first place). Though even there, we might expect a degree of diversity when it comes to particulars: Some feel concern for their loved ones more strongly than aggression or fear towards the perceived threat: some have a stronger emotional focus on the threat than on what they defend. All these differences impact on how the agent deals with the situation morally. Also, of course, any particular emotional reaction may be

---

2 As I shall argue below, we can make a distinction between universalisation across agents and across cases. So a particularist might deny that all agents need endorse a moral judgement for it to be valid, as when Kant says we should only make laws that we could wish all to act on. Or they might deny that we can make valid generalisations that apply across cases, as when someone might claim that causing harm is always bad. While writers who identify as particularists (e.g. Blum, Dancy, McNaughton) often tend to hold both views, they are logically distinct and need not entail each other.
more or less in character. While we would expect that a person show a reasonably stable emotional character across their life-time, some emotional reactions might surprise and mystify us. And some emotional reactions might be both out of character and take strange objects.

Whether such randomness is common, in the psychology of individual agents or in the psychology of humanity as a species, is a largely empirical issue and not my main concern here. Radical randomness, where neither the character of the object nor of the subject explains the emotion, seems likely to be uncommon - but I shall leave that argument for now. Partly for the sake of simplicity, I shall be concerned here with the kind of diversity that is about differences in how individuals are constituted, in a fairly consistent way, across their life-span. But I shall also be concerned with this kind of ongoing distinctiveness of individual agents here for two reasons that are more specific to the thesis topic.

First, as I have suggested at various points in the preceding chapters, emotions are often associated with particularity, and reason with the transcendence of particularity. I have given some reasons to think that emotion need not be indelibly associated with the particular: for one, as I have argued in Chapter Four, agents are capable of emotional identification with points of view other and more general than the one they contingently happen to occupy. I have also argued (in Chapter Two) that particularity need not be normatively problematic in all cases: There are even meta-ethical positions that privilege particular perspectives over general ones. And some meta-ethical approaches that acknowledge the force of more general perspectives also insist on a place for particular ones (Nagel 1986, Crisp 1996, Korsgaard 1996). So we may be able, at least to a degree, to get issues about emotion apart from ones about particularity. But we may also be able to mount a defence of particularity. And insofar as emotion is thought likely to reflect particularity, such a defence could then be used in favour of emotion. I have also argued, in Chapter Two and again in Chapter Five, against the view that particularity is necessarily heteronomous - which brings me to the next point.

Secondly, as I suggested in Chapter Two, the particularity of an agent may

---

3 On character and emotion, see also e.g. Morton (1980), Watson (1987).
have a constitutive role to play in enabling their agency. I have argued that emotional experience, particularly in formative stages of the agent's development, plays a constitutive role in character. I have also argued that emotion subsequently tends to reflect and reinforce such character. And I have argued that a sense of self, in the thick as well as the thin sense, has a significant part to play in constituting the agent as a coherent unit of agency. For developing a consistent sense over time of what one is, what is important to one, what one values, is crucial in informing and motivating choices. Or at least it will be, if we do not want people to be what Frankfurt (1981, 1987) calls wantons, creatures whose actions are more like (random) events than manifestations of agency.

And the individual agent is not only numerically but also qualitatively distinct from other agents. This may be a contingent fact: we can perhaps imagine, albeit with difficulty, a world of agents who were only numerically and not qualitatively distinct. And we can imagine fictional agents who embody Hume's general point of view, Rawls' original position, or Nagel's view from nowhere, or some ideal Kantian or utilitarian universal agent. For such entities, it may be possible to act consistently on behalf of something other than one's own contingent particular self. It may even be possible for such creatures consistently to enjoy goods that are not goods for them. So they might be able to consistently rejoice in the greatest happiness of the greatest possible number even where such a goal is pursued at the expense of their own particular interests.

But these creatures are, precisely, imagined creatures - fabulous monsters. However useful such creatures of imagination may be to counter the vagaries of particularity, they, like the lower-case stoic, are useful as correctives only when used in limited ways. Granted, someone who was never capable of acting on behalf of, and rejoicing in good from, some point of view other than or more general than their own contingent one, would hardly qualify as a moral agent in any normal, non-egoist sense. They might be capable of acting for, and enjoying on behalf of, such

---

4 Frankfurt's wantons are of course distinguished by their incapacity to form effective higher-level preferences, but I take it that extensionally at least they will also have the traits I just attributed to them.
others as they include in an extended version of their basic self-interestedness. But conversely, someone whose moral attitudes were all on behalf of some abstract notion of humanity and never on behalf of actual concrete individuals might also be morally flawed. For such a point of view is a useful abstraction to bring to bear as a corrective when we cannot see the wood for the trees: But it seems equally undesirable that we should be unable to see the trees for the wood.

It is often thought typical of emotion that it can only be extended to a few concrete individuals. It seems intuitively plausible, for instance, that there are real-world limits (though possibly wide and flexible ones) to how many people we can genuinely love. (Cf. for instance Tov-Ruach, 1980). However, I argued in Chapters Four and Five that we are capable of imaginative emotional identification, of feeling from points of view that we do not actually occupy, and that no actually existing agent may occupy. So it seems at least theoretically possible that we could adopt a transcendent point of view through emotion as well as through reason. So, for instance, theologians may speak of divine love as infinite. But, for us, being finite creatures, it might not be desirable or possible that we should adopt such a view - or at least not that we should adopt it consistently. At the very least, in doing so we may lose sight of our own particular point of view - become more saintly at the cost of becoming less human. Failure of transcendence may not be a necessary feature of emotions - but neither need it always be a bad feature when it occurs.

Taking all this as a starting point, I will now offer an argument for diversity in the character and thick sense of self of particular agents to be at least permissible, perhaps even required, within the moral sphere. But I also argue that making this case involves potentially stumbling over a number of antinomies in our meta-ethical

---

5 Some writers of course view such 'extended egoism' as natural and relatively unexceptionable. As Adam Smith portrays the natural bent of sympathies: "After himself, the members of his own family, those who usually live in the same house with him, his parents, his children, his brothers and sisters, are naturally the objects of his warmest affections. They are naturally and usually the persons upon whose happiness or misery his conduct must have the greatest influence. He is more habituated to sympathise with them. He knows better how every thing is likely to affect them, and his sympathy with them is more precise and determinate, than it can be with the greater part of other people. It approaches nearer, in short, to what he feels for himself." (1759, Part Six, Chapter One). However, his notion of the virtuous person is of course someone who is capable of extending sympathy in an altogether more general and considered way.
thinking. Specifically, I argue that making a case for diversity may involve putting considerable strain both on any universalist notion of a community of moral judgement, and on any monist (as distinct from pluralist) notion of moral reality. However, as I have argued throughout, such antinomies may be due to unresolved, and possibly unresolvable, tensions in our meta-ethical thinking, rather than being problems about our emotional natures. It may also be an important consequence of embracing diversity that we will need to accept some degree of luck in how well matched particular combinations of character and situation are - and that too might give some pause for thought. Then again, total flexibility of character to fit any circumstance may be a practically and psychologically incoherent goal.

3) Does morality impose uniformity across agents?

That non-moral differences could be permitted seems uncontroversial, as long as we allow that there are areas of life that are morally indifferent. If one person prefers to spend spare time on outdoor sports and another on reading or in front of the TV, for instance, that hardly seems to be a problem. Except, perhaps, to those who would find either activity too frivolous a way to spend time when there are wrongs to right in the world. (See Kagan 1989, Unger 1996: For counter-arguments, see e.g. Wolf 1997). Of course, it might be hard to find morally insignificant differences in other respects too, since differences in non-moral areas could impinge on moral agency as well. For instance, what people are afraid of, and how well they manage fear, affects what actions they are capable of carrying out (at least without special courage-gathering efforts), which in turn affects their capacity for doing right. And so, of course, do their particular attachments and related commitments.

If, on the other hand, two people are different in moral respects, does that raise problems? I do not have in mind here quantitative differences, as when one agent spends more time, effort and thought on doing good, or more often sees what the right thing to do is and acts on it, than another. In such cases, one agent is clearly more or less morally able than another. Rather, what I want to examine is a problem
about qualitative differences. Suppose we have two people between whom it would be hard to decide who is morally better, and that they differ at least sometimes in the moral judgements they make, or in the way they make them, what they pay attention to and give weight to. Say, to take a hackneyed example, that one generally has a stronger sense of justice and the other a greater capacity for sympathetic understanding, and that they each tend to judge problem cases, and perhaps even non-problem cases, according to those general inclinations. What are we to make of such differences? Should we be concerned about them, and what consequences should there be for how we understand moral agency? I do not want to say that sense of justice and sympathetic understanding are always and necessarily going to diverge. But it seems quite possible that they may do so in particular instances, and that they may be sufficiently distinct and unconnected traits that an agent could have one without necessarily having the other.

In what follows, I will look at some ways in which qualitative differences could be thought problematic, and how such differences would be understood. I will also look at some reasons why we might want to allow room for differences. Finally, I will try to show some ways in which differences can be argued to be at least permissible, and what concerns this might in turn raise.

**Uniformity - some usual (and one not so usual) suspects**

First of all, does anyone actually think such differences should be a source of moral concern? Well, here are some reasons to think so:

Firstly, there is the still influential if much criticised Kantian insistence on universalisability, usually motivated by the implicit or explicit view that particulars distract from proper ethical reasoning. Correct ethical reasoning will be the same for all agents, and operate by the same rules across all cases. In other words, these rules apply uniformly across cases and agents: If lying is wrong, it is always wrong, and for everyone. And any moral agent who judges right will agree on this. While this is probably too simplistic to be fair to Kant or those influenced by him, it will serve for present purposes. (See also my previous discussion of universalisation in Chapter Two).
Secondly, there is a utilitarian pressure towards uniformity. Here, the main consideration for ethical reasoning is what outcome would be best overall. It is assumed that agents who reason correctly (morally speaking) will agree on this point, regardless of their other differences, and seek to bring about this overall greatest possible good. The principle of pursuing the greatest possible good applies uniformly across all cases and agents. So that although the kind of outcomes that follow from applying the principle may be diverse, the fundamental description under which they are sought (and under which they qualify as good) never varies.

Thirdly, and perhaps more surprisingly, there is Aristotle, who in The Nicomachean Ethics states that "men are good in but one way, but bad in many" (1106b35). This is preceded by the claim that “evil belongs to the class of the unlimited [...] and good to that of the limited”. Without wanting to get too involved in contentious exegesis, how can we understand this argument? I think we can reconstruct the line of reasoning, plausibly if not definitively, as follows.

First, virtue is, for Aristotle, defined crucially in terms of the capacity for correct perception of the morally relevant features of any case before the agent - followed, obviously, by correct action. Emotion and desire play a crucial role in such perception (see e.g. NE 1.13). But getting the emotion right seems to be a matter of perceiving the object of emotion right, rather than allowing license for more idiosyncratic emotional reaction.

Secondly, there is the doctrine of the unity of virtues. One possible reading of this is as the claim that any truly virtuous person will have all virtues, not some more or less random selection that may differ from that of other virtuous persons. Or, following McDowell in “Virtue and Reason” (1998: Chapter Three), we might think of virtue as a central capacity which expresses itself in particular virtues - courage, honesty, compassion, a sense of justice - as and when the situation requires. In this case this second point may collapse into the preceding one, about virtue as correct perception and reaction. By extension from either reading of the doctrine of the unity of virtues, it is usually thought to follow that a virtue cannot exist as in isolation: you cannot have real courage, compassion, and so on, if you do not possess virtue in some more general sense. No particular virtuous trait has real
existence apart from the central capacity\textsuperscript{6}.

The Aristotle angle may be worth some particular emphasis in so far as it is often assumed that virtue ethics can accommodate moral pluralism more easily than deontological or utilitarian ethical approaches can. For those that draw on Aristotle for their virtue ethics, it might be necessary to re-examine some assumptions in Aristotle before this approach will do better for pluralism than the competition.

\textbf{Some reasons we might want to allow diversity}

Trying to side-step difficult issues about cultural climes here, officially at least our current cultural clime places high value on individuality and diversity\textsuperscript{7}. These may not in themselves seem to be obviously moral features to promote. But if they are valued as forms or constituents of human flourishing, then it seems that they are, at least \textit{prima facie}, goods that morality ought to promote.

Moreover, and possibly related, we might reasonably be reluctant to see all differences with regard to moral issues as necessarily being a sign of deficiency on someone's part. This view can arguably be held even by those who would reject any facile kind of relativism. As I argue below, in at least some cases the facts of the situation may underdetermine what the correct moral response (judgement, emotion or action) is. This need not mean that anything goes, but it does suggest that sometimes, at least not only one thing goes.

Also, as Strawson (1974c) suggests, we might think that there was more than one way in which it would be good for people to live, but that not all the ways in which it would be good for people to live could be realised by any one agent. Some of them might not be practically or logically possible to combine. There are only so many interests any one person can pursue in a lifetime, only so many good works they can perform. There being different kinds of agents, pursuing different versions of human good, might then be the only way in which different goods could be realised together.

Finally, as suggested earlier, there may be important links between

\textsuperscript{6} But on the doctrine of the mean, see also Urmson (1973).
\textsuperscript{7} There are of course also tangled political philosophy issues here: On diversity, particularly in relation to emotion, as an issue of \textit{social} morality specifically, see e.g. Strawson (1974c), Scruton
distinguishing features of agents and the kind of internal coherence required for adequate moral agency. Capacities for integrity and autonomy are often thought to be connected to distinctive features. Certain ways of looking at the world, types of responses, and the commitments that follow from these, may be not only defining of an individual but also play a crucial (if not necessarily exclusive) role in enabling them to build the internal coherence necessary to be an efficient agent. Someone with no preferences would be adrift in the world – and also rather hard for other agents to interact with. And if such features are required for efficient agency, then it seems probable that they will also be required for efficient moral agency.

To support these points in favour of pluralism, though, it will be necessary to re-examine some of the assumptions that underpin pressures towards uniformity. Picking up the thread from Kant, the utilitarians and Aristotle again, then, the next section will examine some sources of such pressures.

Some general pressures towards uniformity

Firstly, we have what I shall call the direction of fit-assumption: the aptness of moral judgements is generally assumed to be given by their fitting the facts of the case before the agent. The exact ontology of moral facts, and their relation to non-moral facts, may be obscure, and is certainly contentious. But the direction of fit for correct moral judgement, as for other forms of judgement, is often thought to be from world to agent. The agent's input into correct moral judgement may in a sense be a merely passive recognition – rare as it might be for such recognition to be entirely accurate. This view need not involve a commitment to any strong realist position: it may be merely a buffer against acceptance of random projection. But then again we might think that in avoiding projection like this, we have gone too far the other way: Now there is no input from the agent, rather than only input from the agent.

A related, second, assumption: Correct moral judgement will be that in which competent agents concur. I shall call this the agreement-assumption. So, if agents disagree about a case, that must be due to some difference in moral capability

between the agents, for instance a lack of ability to abstract from the agent's own part in the moral issue at hand. Partisanship and self-interest, for instance, may be seen as powerful distorting influences here. But apart from such motivated cases of distortion, characteristic habits of perception, feeling, motivation or action might all be thought to exert biasing influences.

Thirdly, the notion of morality as a rule-governed practice - the rule-assumption - also seems to impose natural constraints on individualism. If morality is rule-governed, personal differences in judgement will be constrained by the rules. This argument, as indicated, applies particularly to individual differences. If the private language argument holds, it seems that we cannot have a moral constituency of one. This does not rule out the possibility that groups might differ in the rules they apply: That possibility is however constrained if we accept the previous points. It is also worth noting that the direction of fit-assumption and the rule-assumption, though independent, are compatible and mutually supportive. Rules about reasoning may be justified in terms of their making it more likely the agent will pick out the relevant features of whatever case they are considering.

Both the direction-of-fit-assumption and the rule-assumption point towards an understanding of correct verdicts on cases as being ones that morally capable agents will agree on, and so towards the agreement-assumption. The rule-assumption also points towards a requirement for consistency across cases that can be held regardless of whether one subscribes to the direction-of-fit-assumption or not. And if morality is merely an interpersonal construct, all the more reason to insist on agreement, at least about basic rules.

If we accept these assumptions, is there much room left for difference, individual or otherwise, to be compatible with adequate moral agency? The direction-of-fit-assumption leaves relatively little room for individual or indeed any input from the agent other than a relatively passive recognition. The agreement-assumption suggests any difference in opinion between agents would be down to someone having blundered. And the rule-assumption suggests the blunder to have consisted in a failure to abide by rules that bind all agents insofar as they aim to be good moral agents. The overall conclusion seems to be that difference can at most be
permitted in areas of moral indifference, if such exist.

4) Some possible ways to counter pressures towards conformity

1) Deny complete determination

In at least some cases the morally relevant facts may underdetermine the correct overall moral response to the case, leaving more than one type of response to the case legitimate given the underdetermination. Our moral principles may yield uncertain or contradictory results when applied to the facts before us. Moral dilemmas will, I take it, be examples of cases that are underdetermined by principles; there may be other kinds.

Such denial of complete determination will still be compatible with the view that some responses or kinds of responses will not be morally acceptable possibilities: it does not mean that anything goes. Nor does it mean that some alternatives cannot be ranked higher or lower than others - but it does mean that there is more than one alternative, and that there will (if the case is genuinely underdetermined) be two or more between which there is nothing definite to choose.

There are a number of different ways in which such underdetermination may be understood to obtain: I will not discuss that issue in depth here. One simple way of getting to underdetermination - or at least local instances of it - might be to take it that there are several different, possibly conflicting, and not necessarily internally rankable sources of value.

So Nagel (1979d), lists five “fundamental types of value that give rise to basic conflict”: specific obligations to other people, institutions or causes; constraints on actions deriving from basic human rights; utility; endeavours of intrinsic value, such as scientific discovery or high artistic achievement; and commitment to one’s own personal projects. These may conflict, and there does not seem to be some all-purpose indefeasible way of ranking them. And to these we might add further kinds of conflicting value, as Nagel himself does in claiming that agent-relative and agent-neutral points of view may clash, but endorsing both as useful and necessary ways of looking at the world and discovering value (Nagel
1979h, 1986).

Underdetermination may apply both at the level of specific situations, and at the level of broader patterns of choice across situations, such as lifestyle choices, or choices about what kinds of moral causes to promote most strongly - given that any individual can only get so much moral work done in a lifetime.

Any otherwise inoffensive distinguishing characteristic of the agent may then have a valid, even a positive contribution to make if it tips the balance one way or another between the equally ranked alternatives. Though of course this holds only on the further assumption that choosing is better than not choosing - and still leaves worries about whether weight needs to be given to the considerations supporting the alternatives that are not chosen.

This line of argument, if workable, opens up the possibility of there being areas of moral indifference. Not in the sense of non-moral pursuits an agent might want to engage in and that lack moral implications, but in the sense of there being morally legitimate options that there is no clear way to rank by the criteria we usually apply to moral decision-making.

The crucial point here is that a case in which two or more responses are equally valid would pose a problem for any of the three forms of pressures towards uniformity, at least if these principles are strictly interpreted. For in such a case, there is room for individual input - albeit within constraints - in choosing one response over another.

Underdetermination can of course be understood to hold only at the epistemic level. It is thinkable that there is in all cases some one right answer, but that we simply cannot, being creatures of finite capacities, access this answer. However, if we aim to deal with the problems of human virtuous agents, not with angelic or godlike ones, then I think we can plausibly assume that underdetermined cases (at least at the epistemic level) do exist. Nor does underdetermination at the ontological level seem obviously impossible.

8 On incommensurability of values, see also Chappell (2000). The possibility that there may be incommensurable values may also be used to try to vindicate the possible rationality of emotional ambivalence (for a discussion of some of the issues here, see Harris, 2001).
2) Deny indiscriminate application

Even if the case should be fully determined, it is not clear that this means what should be done, or even what reaction should be had, applies in the same way to all agents. Some may be thought to have a particular obligation given their role in a situation or their relations to other agents or entities involved. In some cases, mere presence may be argued to constitute an obligation to act in response to the situation. Williams' famous example of Jim and the Indians (in Smart and Williams, 1973) being arguably a case in point. Conversely, some may be thought to have no such obligations in a particular case, or to have weaker obligations than others.

There might even be empty obligations, that hold for no-one. For instance, if a child is left with no living relatives or near significant others capable of looking after it, we would hardly want this situation to continue. But then again it seems there are no obligations on individual agents unrelated to the child to take care of it themselves. Unease about such cases may be part of the motivation for creating artificial agents, such as welfare agencies, which can do moral work that needs to be done but that no-one has an obvious obligation to do.

This argument does not by itself give grounds for different agents to view the case in itself differently. By distinguishing judging a case from acting in response to it based on a further judgement about how it applies to oneself, this argument does however provide some more space for diversity to be compatible with adequate moral agency than implied above. And a moral agent who is primarily practically minded may tend to think of, and even notice, the situation first in terms of how it applies to himself specifically as a moral agent.

Furthermore, the relations of the agent to the case may themselves be morally underdetermined. There may be no exact truth, or no accessible exact truth, as to what obligations follow from standing in any particular relation to a case. And that last point is in turn further complicated by the fact that agents often stand in complex relations to any situation through different roles, and through the impact of their roles in other situations, some of which may not be directly related to the current one.

It is worth pointing out that this kind of contextual detail is something that
emotion is often thought particularly responsive to and reflective of. (On this point see especially Blum, 1980a, 1994). Emotion may also itself be constitutive of obligation, where it bonds the agent to others. There may also be relations that an agent stands in to situations in virtue of their individual character and the abilities that their character traits confer – a point that, if true, would increase vulnerability to moral luck. This point may also possibly have the unwelcome effect, for the more morally capable agent, of increasing obligation on them relative to less capable agents. Though conversely, agents in general might be thought to be under some obligation to transcend the limitations set by givens of character and circumstances. We would hardly want contingencies of someone’s original context of moral learning to entirely exculpate them from moral obligation if they should find unexpectedly find themselves in different circumstances.

As the above suggests, this is a point on which we might feel considerable ambivalence. On the one hand, we might not want to leave the door open too wide to moral luck, and issues of application seem particularly likely to have this effect. On the other hand, we might want to acknowledge that such vulnerability seems to be a fact. Someone could lead a fairly blameless life and then find themselves out of their moral depth when suddenly in an unfamiliar context - as the Milgram experiments for instance suggest. As far as particular obligations go, we might be particularly concerned to avoid the possibility of empty obligations - but still doubt whether we can always rustle up a suitable form of collective agency to shift such obligations onto.

But at the other extreme, we would hardly want to find each individual agent effectively saddled, Kagan-style, with the obligation to promote the greatest happiness of the greatest possible number to the full extent of their ability at all times, regardless of the cost to their chances of individual flourishing. But between these extremes, it seems there should be room for people to disclaim some responsibilities and acknowledge others, and also room for a degree moral specialisation not to be altogether a bad thing.
3) Consistency between agents versus between cases

Thirdly, we might think that one implication of underdetermination is that the requirement of agreement or consistency between agents comes apart from the requirement for consistency across cases. This points to some further implications of the possibility of underdetermination to be teased out here. For if cases can be genuinely morally underdetermined, so that there are two or more alternatives that there is genuinely nothing to choose between, then interpersonal agreement about the choice of any one option over others cannot straightforwardly be based on the content of the agreement as validating the choice of that alternative over other, morally equal ones. And in that case, interpersonal agreement does not seem to provide much better a guide to choice in these cases than more individual reasons. We might still insist that interpersonal agreement, or some (other) version of universalisability, would constrain what the morally equal options are. But it cannot, if the case is genuinely underdetermined, tell an agent which option to choose from among the permitted ones.

Underdetermination also presents a problem for the notion of morality as a rule-governed activity, insofar as if the case in hand is morally underdetermined by the (morally relevant) facts, that implies that it is also underdetermined by any rules based on relevant similarity between cases - assuming, of course, that relevant similarity is a meaningful notion in the first place. It may still be thought that we would need some kind of consistency, for instance in the form of background attitudes being constant across cases, to avoid mere randomness about the content of judgements, but it seems difficult to articulate properly what this kind of consistency would mean. It cannot, it seems, be consistency based on the facts before the agent. And if genuine moral dilemmas can exist, appeals to sensibility rather than rules may not be altogether helpful, either. At least not if the sensibility in question is of the kind envisioned by McDowell in unity-of-virtues mode: for if the facts of the case genuinely underdetermine what the correct moral response is, sensibility qua object-sensitivity can hardly be expected to solve the problem.

In such cases, it seems the choice may go to the individual agent, and perhaps in such cases the distinctive character of the agent presents (as a basis for
choice) the last hope of avoiding some greater randomness. The argument in this section may then have established some room for difference within the moral sphere. How significant this is will largely depend on how prevalent one thinks morally underdetermined situations are – and the room for difference may have to come at some cost to the extent to which moral decisions can be interpersonally intelligible and based on shared conceptions of rules.

Moreover, this way of making room for differences raises the question of the moral status of the agent's character - and whether it can genuinely do work that facts, rules and agreements cannot. It also raises the problem mentioned at the outset: If two agents are morally equal but also qualitatively different in their moral capacities, is this difference morally indifferent, or should it concern us?

5) Some worries about the virtuous agent

Suppose we consider again the notion of virtue as the capacity for reliably correct moral judgement. The Aristotelian virtuous agent, for instance, will be angry, merciful, kind, and so on, in the right way, to the right people and at the right time, and for the right purpose. Morally speaking, he (pronoun concession to Aristotle) fits into the world just right. On the other hand, it is not clear that character here is doing much more than providing a stable tendency, rare and valuable as that might be, to get things right by some standard that is not itself set by that character. And problems about such standards, such as under-determination, have already been raised.

How does a virtuous agent decide between the permissible alternatives when faced with an underdetermined case? The decision can either be random, or part of some pattern of the agent's. If the arguments above work, it seems that we lack other plausible sources of patterning at this point. Referring the case to intersubjective agreement, for instance, may not be a workable option, except perhaps in a "let's take a vote" way. In either case, uniformity with other virtuous agents is threatened. Unless we assume that faced with underdetermined situations virtuous agents will all pick the same option, that is - and since there is no more reason, on the merits of the
case, to pick one option rather than another, this seems improbable.

If such decisions are made randomly, there is a further problem: Unless the agent is extremely lucky and comes across no or at least very few - underdetermined situations, randomness of choices would undermine his claim to form a sufficiently coherent entity *qua* agent.

If the choice conforms to a pattern, does this mean that the agent's way of being virtuous is after all different from that of other virtuous agents? Suppose he finds himself, to take a fairly trivial example, in a case where honesty and kindness pull in opposite directions without being clearly rankable. If his decision to go with one or the other is part of a pattern, then we may assume the pattern will rear its head also when the agent comes across cases that are not underdetermined.

For character traits, once present, have strong self-reinforcing tendencies. As Rorty (1988) argues, character traits do not just predispose us to react to and interpret what is before us in characteristic ways, even beyond what is objectively warranted. They also dispose us to gravitate towards, even bring about, situations that elicit the exercise of the trait. And where the agent sees themselves as possessing a particular kind of trait, dispositions sometimes appear in second-order form: the agent may deliberately, even consciously, gravitate towards and bring about occasions to exercise the self-ascribed trait.\(^9\)

It may also be that acting in a particular way in some underdetermined situation will by itself establish a pattern, if the underdetermined situation is one that plays a formative role in the agent's development.

If there is a pattern that is the agent's own, then his moral decisions may no longer be determined purely on the merits of what is before him: his own particular traits will tend to sway the decision. If on the other hand there is no pattern that is the agent's own, there seems to be little obvious option other than randomness for the agent in underdetermined situations.

If we think that forsaking a notion of virtue as being the same for all agents is more tolerable than facing underdetermined situations with random choices, then we also have a case for allowing agents to develop distinctive patterns - and for

\(^9\) For a rather curious example, see Wollheim's "Pale criminals" in Wollheim (1993).
these patterns to legitimately direct agency, at least in underdetermined situations.

Since the options ranked are otherwise equal, it seems undesirable, or at least not clearly desirable, that all agents should make the same choice of pattern. Otherwise, we might end up, for instance, with all and only such agents as *ceteris paribus* value truthfulness over other principles. In other words, not only patterns but also a variety of patterns may be possible to justify. Such patterns may have a particularly significant role to play in big decisions of the life choice type, e.g. what kind of work to do, and what kind of people to let into one’s life. For these kinds of questions are typically underdetermined at least at the epistemic level, due to the uncertain nature of options and likely outcomes. They may also be particularly vulnerable to conflicting sources of value. But that is not to say that they are not, at least in large part, also moral issues.

**Costs of allowing patterns**

However, there are costs to patterns - some have already been indicated above. Patterns may cause not only a motivational bias to act in conformity with the pattern even in underdetermined cases, but also a cognitive bias to view cases in conformity with the pattern. Agents’ having different patterns may, especially given these cognitive and motivational biases, constitute a significant obstacle to interpersonal agreement on moral assessments, which will be a further problem if patterns have a tendency to spill over onto cases that are not underdetermined. Overcoming the problem of the biases at least to the point of making communication possible may require a considerable, and perhaps not consistently sustainable, level of critical self-awareness on the part of any agent.

It will also require other agents to be willing and able to recognise and understand such idiolectic tendencies. And recognising and understanding may be clearly distinct skills here. It takes relatively little skill and knowledge to spot a distinctive tendency in actions and attitudes. It takes rather more experience and interpretative effort to understand the idiolect, penetrate its meaning, and be able to read the person right. For what one person thinks of as showing respect for their own and other people’s boundaries and privacy, another may think of as cold remoteness;
directness may be read as rudeness; concern as intrusiveness; interest as inquisitiveness. Social ideologies of emotion may provide useful basic frameworks for reading here, but may also confuse. Particularly so since agents now more often change their settings, moving across social and national boundaries: in the process, moral language barriers may be encountered.

Part of the appeal of the doctrine of the unity of virtues is that it would avoid such problems - as long as there are no significant underdetermined cases to scupper the agent’s progress. If there is only one way in which we go right, if virtuous agents share the same sensibility or rules, the potential for misunderstandings between virtuous agents is at the least considerably diminished. Virtue so understood would also not run the risk, which patterns do, that the pattern one has developed need not be one particularly suited to making a decision in the particular underdetermined situations one comes across. Assuming, of course, that genuinely underdetermined situations are possible - some might deny this.

Nor is it always obvious which patterns may be in conflict in a situation. An agent might for instance have both a commitment to truth-telling and an inclination to spare people’s feelings, and so still have the dilemma as strongly. Virtue as a central unitary ability responsive to the merits of cases would not be skewed in the ways that a more distinctively patterned form of virtue is. Which is presumably why Aristotle seems to have thought such more distinctively patterned forms were at best imperfect virtues and perhaps not true virtues at all.

**The unity of virtue and the unity of good**

However, it seems that the doctrine of the unity of the virtues can only be sustained if we can give a relatively internally coherent and unitary account of the moral good. If we allow any kind of fragmentation into the notion of moral good, so that there may be significantly distinct and possibly competing sources of value, the doctrine of the unity of the virtues, and by extension the sameness of virtuous agents, may become untenable.

In this case, it seems that we may want to allow that there can be different ways of being good, and that the difference may be qualitative rather than
quantitative. So we might allow that agents could be differently but equally good. In other words, we may have made a place for diversity in moral agency.

But this option comes at some cost of its own. For instance, the matching-up of situation and individual character may be crucially vulnerable to moral luck. So while we might want to argue, as Flanagan (1991) does, that virtue is multiply realisable, we might also worry about whether its particular realisation in any particular agent will be well matched with such moral challenges as come their way. And this is not a trivial point. If morally good character is a good that is enhanced by executive virtues such as consistency and integrity, then it seems we may have to either forgo such enhancement or else risk having characters too rigid – or just plain unsuited - to meet the demands of particular situations.

Moreover, if particular patterns are thought to be defining of agents, there may be a case for viewing the particular character of the agent as a basis for judging that agent as standing - or not - in an obligating relation to behave in a particular way in a particular situation. Does character-type make one liable or not liable to certain obligations? This seems to raise issues similar to the more general one of whether talents obligate - and whether lack of a particular talent excuses from effort. Moreover, in the absence of anyone particularly suited to do whatever moral task is required, the question of obligation without particular relation to the case, by character or otherwise, arises with considerable force. In other words, allowing patterns may involve allowing the possibility of inequalities in moral workload – and the possibility of empty obligations.

How seriously this concerns us partly depends on where we put the emphasis, and of course on how perfectionist we get. If, for instance, our main priority is to have as many good agents as possible, regardless of whether they are the same or not, we may not be too worried about exact match between character and circumstance in all cases. On the other hand, if our first priority is that the right kind of things get done, then we may have serious worries if it seems that reaching this objective is vulnerable to luck in the psychologies of the available agents. All very well for them to be sustaining internal coherence in their particular take on the virtues, but if we get, to use some of Flanagan’s examples of candidates for moral
sainthood, a Mother Theresa when we need a Mahatma Gandhi, or vice versa, we may worry about how much value exactly is in diversity. And then we might think that if the agents available can be reshaped to fit the task, any sacrifice of their internal coherence and distinctive character is one we can afford.

And of course, allowing for such agent-relativity as goes with diversity may limit the interpersonal intelligibility of moral decisions, and the degree to which they can gain validity by appeal to shared reasons. And this means that moral pluralism raises considerable challenges to agents' capacities for tolerance, understanding and respect in the face of difference.

Making room for differences, then, seems to come at a cost of allowing for considerable vulnerability to moral luck and a cost to the shared nature of morality, even before the question of making room for people's non-moral features and pursuits has been considered. And it seems we cannot simply assume that even the virtuous agent will morally speaking fit neatly into the world.

A more flexible notion of character?

Of course, as suggested above, we may want to make concessions on the agency-side of the equation instead. For instance, we might relax the insistence on internal coherence in agency considerably. A too monolithic concept of character may be implausible both empirically, as a way of understanding how people actually are constituted as agents, and normatively, as a standard to which agents ought to conform. The normative standards in question might range from purely functional concerns about efficiency to more substantive evaluative concerns. It may be that a more plastic notion of character will undercut worries that the capacity for efficient agency will be lost if intrapersonal consistency across cases is relaxed.

And on the other side of the equation: It may be, though our moral theories at the moment do not seem up to the task, that we will at some point find a plausible method of dealing with problem cases in a consistent manner, so that reliance on patterns of the agent may in any case become a moot point - someday. Though this would still leave worries that internal consistency on the part of the agent might be a good in itself; one that our meta-ethical theories and moral practices should at least
ceteris paribus seek to promote or at least allow, even if such consistency is not instrumentally required for good moral agency, and may sometimes work against it.

6) Emotion, difference and moral agency

In conclusion, then it seems worth pointing out that insofar as we have been able to meet worries about the subjectivity of emotion, it has largely been by bringing it within the kinds of rule-models, objective or intersubjective, that power pressures towards conformity. Thus it seems that we could make emotion more meta-ethically respectable by arguing that it reflects salient features of its objects, in ways that are at least intelligible if not predictable. Or by arguing for emotion's conforming more than is usually thought to standards of intelligibility and appropriateness, and for the intersubjective nature of these standards. The remainder worry was about the degree of idiosyncrasy that the particular emotions and the emotional characters of particular agents still display within these broad constraints.

However, I have also argued for the intelligibility, rationality and value to moral agency of emotion in terms of how emotion reflects the character of the agent that feels it. Emotion is an invaluable and irreplaceable, if not always pleasant, means to self-knowledge, through what it reveals about us. And I have argued that while emotion may reflect, and even be partly constitutive of, various kinds of failures of autonomy, it may also reflect us in our autonomous aspects, and play a similar part-constitutive role in those. I have also argued that emotion plays a constitutive role in creating and shaping our sense of self. I have argued that emotion crucially informs and motivates moral agency. So another part of making emotion respectable is to point to its connection to who we are, what we are like.

But I have also argued, in this chapter and earlier, that the standards by which emotion has often been found wanting, and by which I have argued here that it may in fact largely be vindicated, are themselves in some tension, separately and together. So it seems that while emotion is less normatively troubling than its critics would have it, the normative standards applied in attacking emotion may themselves be trouble-prone. And insofar as emotion is vindicated with respect to these
standards, the same problems that those standards are prone to will apply to emotion. Which is not to say that emotion would look too good without some vindication with respect to these normative standards, or that such vindication is not based on rather more descriptively plausible accounts of emotion than those offered by at least some of emotion’s critics. But emotion itself does not seem to be the problem: rather, it is used as a McGuffin for unresolved problems in our meta-ethics and our moral psychology. So the defence of emotion stands - with such qualifications as have turned out necessary along the way. Getting rid of emotion will not solve such problems as we face over moral agency. And emotion may be a vitally necessary help towards solutions.
Conclusion

There are three main points I want to emphasise in this final summary.

The first is about general features of emotion: I have argued that emotion as such need not be a problem for moral agency, and is in fact crucial to it.

The second is about the particular form emotion takes in human agents. I have argued that where emotion is a problem for moral agency, this is largely because of the particular form it takes, rather than being down to emotion as such. And the particular forms emotions take are due to factors external to emotion.

This leads to the third point: Insofar as the problems are not really about emotion, and have their sources outside it, emotion is being used as a McGuffin for unresolved issues about moral agency.

On the first point, I have argued that emotion is best understood as an affective form of cognitive intentional awareness. Emotion has important similarities to and links with beliefs, desires and moods, but is distinct from all of these. In its dispositional form, it also resembles and may shade into more consistent traits of character or temperament. Emotion also plays a formative role in establishing such traits, and tends to reflect them once they are established. So emotion has a distinct but hub-like place among our mental kinds.

Emotion, so understood, need not be a problem for moral agency. More specifically, it need not be liable to the kinds of worries about loss of capacity for objectivity, rationality and autonomy that have traditionally been levelled at it. Emphasis on the cognitive element in emotion goes far towards meeting such worries. But if emotion is not to be made meta-ethically redundant, we need to resist attempts to assimilate emotion too closely to other cognitive kinds such as beliefs or judgements. Such assimilation is also descriptively implausible. It is more plausible - and more meta-ethically useful to a defence of emotion - to emphasise the *sui generis* nature of emotion-cognition, and the crucial role played in it by feeling.

Beyond making emotion less of a problem, viewing emotion in this way also explains how emotion may in fact play an indispensable role in informing and
motivating moral agency. On a general level, emotion, being affective, is an engaged and interested mode of awareness, and as such plays a crucial motivating role. Through hedonic tone, emotion informs as well as motivates, alerting us to what is good or bad for us. And most crucially for moral agency, emotion plays an indispensable role in underpinning participant reactive attitudes. These attitudes in turn are conceptually crucial for recognition of oneself and others as moral agents and moral subjects. Detaching from these attitudes may be appropriate and useful in particular instances. But global detachment - what I have called lower-case stoicism - is internally incoherent as a strategy for moral agency, and leads to a peculiar form of moral solipsism.

Emotion as such, then, does not seem to be the problem, or dispensable. Rather, moving on to the second point, the problem is with the particular form emotion takes in human agents. I have argued that emotions reflect what those who feel them are like as well as those things and persons that the emotions are felt towards. So our emotions reflect human nature and the human situation. If either of those were different, our emotions would be different. Specifically, I have suggested that failures of objectivity, rationality and autonomy may be contingent on our nature and situation and not due to emotion as such. Emotion may tend to reinforce such flaws, once present, but they are not inherent to emotion. Rather, the problems are due to basic features of how we are constituted - our individual separateness, our differences, our vulnerability, our interdependence, and the finite nature of our capacities, including our reasoning capacities.

But it is also part of our human nature that we possess capacities for creative self-determination. There are a number of ways in which we are capable of controlling our emotions and resisting emotional temptations - through argument, through deliberate control of behaviour, through adopting different perspectives, and through ideologies of emotion. We also have, as the stoic insists, the capacity for detachment. And clearly, these capacities are crucial for our capacities for good moral agency. However, I have also suggested that we need to exercise some degree of caution about how we use these capacities. For they, as well as emotions, may go wrong or too far. We may be flawed or misled in our reasoning. Or we may, like the
stolic, take detachment too far, and by over-application turn a useful capacity for
detachment into a debilitating incapacity for engagement. It is also notable that our
capacities for sympathy and empathy, which provide crucial means of controlling
and improving our emotions, themselves depend crucially on our capacity for
feeling. So again, good moral agency is not all a matter of overcoming emotion.

That there are some worries about how we should apply our capacities for
shaping and controlling emotions also leads to the final point: the McGuffin theme. I
have argued that the normative assumptions that power the objections to emotion are
themselves problematic, separately and together. Subjectivity, in its various senses,
may play a positive role in constituting individual agents as internally coherent
agents. And loss of autonomy is unfortunate, but openness to the influence of givens
and to the points of view of others is also crucially useful to keeping our moral
thinking on track. And some forms of apparent rationality-failure may be
unavoidable, if we concede that there can be different and conflicting kinds of moral
reasons that we may need to be sensitive to.

But these are not necessarily problems about emotion. Rather, emotion has
become a focal point for various stubborn meta-ethical problems, such as the place of
agent-relative reasons and particular attachments in morality, the value of autonomy,
and the moral status of the givens of our natures and situation. But at least in theory,
our emotions could take rather different forms than they in fact do - be more
impartial, less idiosyncratic, more transcendent of contingencies, less given to
particular attachments and more to generalised altruism, in short more as Kantian or
utilitarian impartialist approaches might wish them.

To effect this change, though, we would have to either change what we are
like or at least act as if such a change had been made. So the entities whose new
improved emotions these would be would not quite be humans as we know them. It
falls beyond the scope of this thesis to mount a fuller defence of human nature. And
any such defence would clearly be qualified by due awareness of our many flaws as a
species and as individuals. But at the very least, given that our capacities for self-
improvement may themselves be flawed, it seems we should exercise some caution
in our attempts to improve on ourselves - our emotional natures included.
BIBLIOGRAPHY AND REFERENCES


Apostoleris, Nicholas H. and James D. Laird 1996: "Emotional self-control and self-perception: Feelings are the solution, not the problem", in Harre and Parrot 1996.


215


Burnyeat, Miles 1980: "Aristotle on learning to be good" in Rorty 1980c.


Cooper, John M. 1996: “Eudaimonism, the appeal to nature, and “moral duty” in Stoicism”, in Engstrom and Whiting 1996.


Epicurus [no date of publication given]: *Principal doctrines*. Sections on emotion reprinted in translation at: http://classics.mit.edu/Epicurus/princdoc.html


A later version of this paper was published in Philosophy, 1998.


James, William 1996 [sic]: “The will to believe”. Reprinted in Feinberg 1996.

James, William. 1884: “What is an Emotion?” Mind 19, 188-204.

Johnson, Lamont 1999, unpubl.: “The primary categories of affectivity: Feelings,
passions, moods and emotions”, conference paper. Provided by J. Jacobs.


Langton, Rae 1994: “Maria von Herbert's Challenge to Kant” at http://www.ed.ac.uk/~ajbird/Introduction_html/KantMaria.html


Lazarus, R.S. 1982: “Thoughts on the relation between emotion and cognition”.
American Psychologist 37, 1019-24.


Blackwell.


Nagel, Thomas 1979g: “What is it like to be a bat?” in Nagel 1979a.


available as postscript file from: ftp://ftp.cs.bham.ac.uk/pub/groups/cog_affect


Rorty, Amelie O. 1980b: "Explaining emotions", in Rorty, 1980a


231


