Empathy, Motive and Morality:  
An Enquiry into the Role of Empathy in Ethics

A Thesis submitted to the University of Manchester for the degree of  
Doctor of Philosophy  
in the Faculty of Humanities

2012

Gayle Impey

School of Social Sciences
Contents

ABSTRACT ................................................................................................................. 6

DECLARATION ........................................................................................................... 7

COPYRIGHT STATEMENT ............................................................................................ 8

1.0 DEDICATION AND ACKNOWLEDGEMENTS ....................................................... 10

1 INTRODUCTION ....................................................................................................... 12

1.1 THESIS OVERVIEW ............................................................................................ 12

1.2 EMPATHY PHENOMENA .................................................................................... 20

1.3 HIGH- AND LOW-LEVEL EMPATHY .................................................................... 26

1.3.1 PROCESS VS. OUTCOME ............................................................................... 27

1.3.2 EMPATHY AND SIMULATION THEORY ..................................................... 28

1.3.3 HIGH-LEVEL EMPATHY ............................................................................... 34

1.4 SUMMARY ............................................................................................................ 39

PART I .......................................................................................................................... 40

2 HUME’S SYMPATHY ................................................................................................ 41

2.1 INTRODUCTION .................................................................................................... 41

2.2 SYMPATHY ........................................................................................................... 43

2.2.1 THE PASSIONS ............................................................................................... 43

2.2.2 THE SYMPATHY MECHANISM ...................................................................... 52

2.3 COMPASSION AND BENEVOLENCE ................................................................ 61

2.3.1 THE PRODUCTION OF COMPASSION AND BENEVOLENCE ...................... 61

2.3.2 IMAGINATIVE SYMPATHY AND COMPASSION ........................................ 73

2.3.3 SYMPATHY, COMPASSION AND MOTIVE ............................................... 78

2.3.4 SUMMARY ...................................................................................................... 83

2.4 MORAL EVALUATION .......................................................................................... 83

2.4.1 SYMPATHY AND THE MORAL SENTIMENTS ............................................ 84

2.4.2 THE COMMON POINT OF VIEW .................................................................... 93

2.4.3 THE PROBLEM OF MOTIVE ......................................................................... 103

2.4.4 SUMMARY ...................................................................................................... 109
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>Conclusion</td>
<td>110</td>
</tr>
<tr>
<td>3</td>
<td>ADAM SMITH'S SYMPATHY</td>
<td>111</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>111</td>
</tr>
<tr>
<td>3.2</td>
<td>Sympathy</td>
<td>115</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Introduction</td>
<td>115</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Outcome vs. Process</td>
<td>116</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Sympathy and Imagination</td>
<td>119</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Sympathy and Simulation</td>
<td>125</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Summary</td>
<td>137</td>
</tr>
<tr>
<td>3.3</td>
<td>Compassion</td>
<td>137</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Aspect (1): Reliability</td>
<td>137</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Aspect (2): Compassion</td>
<td>140</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Summary</td>
<td>143</td>
</tr>
<tr>
<td>3.4</td>
<td>Moral Evaluation</td>
<td>144</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Introduction</td>
<td>146</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Aspect (2): Moral Evaluation</td>
<td>151</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Summary</td>
<td>159</td>
</tr>
<tr>
<td>3.5</td>
<td>Virtue</td>
<td>160</td>
</tr>
<tr>
<td>3.6</td>
<td>Conclusion</td>
<td>163</td>
</tr>
<tr>
<td>4</td>
<td>MICHAEL SLOTE'S EMPATHY</td>
<td>167</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>167</td>
</tr>
<tr>
<td>4.2</td>
<td>Empathy</td>
<td>173</td>
</tr>
<tr>
<td>4.3</td>
<td>Moral Obligation</td>
<td>180</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Introduction</td>
<td>180</td>
</tr>
<tr>
<td>4.3.2</td>
<td>The Bias Problem</td>
<td>185</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Summary</td>
<td>193</td>
</tr>
<tr>
<td>4.4</td>
<td>Moral Evaluation</td>
<td>194</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Introduction</td>
<td>194</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Approval and Disapproval</td>
<td>195</td>
</tr>
<tr>
<td>4.4.3</td>
<td>The Modulation of Empathy</td>
<td>198</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Summary</td>
<td>200</td>
</tr>
<tr>
<td>4.5</td>
<td>Virtue</td>
<td>201</td>
</tr>
<tr>
<td>4.6</td>
<td>Conclusion</td>
<td>202</td>
</tr>
</tbody>
</table>

PART II 205
Abstract

The thesis argues that two ways in which we can consider empathy’s role in ethics are fundamentally flawed because they fail to take into account the myriad ways in which empathy can be affected and influenced by our motivations. I apply what I call ‘the motivation objection’ to these two views. This has three aspects: (1) *reliability*: because empathy can be affected and influenced by our motivations, empathizing does not always lead to the right results; (2) *function*: because it can be affected and influenced by our motivations, empathy is not sufficient for various functions; (3) *circularity*: because it can be affected and influenced by our motivations, empathy cannot be used to define or explain certain aspects of morality in a non-circular way.

The two ways of considering empathy’s role in ethics are what I call constitutive views, according to which empathy in some way constitutes, or is the foundation for, morality, and instrumental views, according to which empathy is of instrumental value in morality. I apply the motivation objection to three constitutive views, two historical and one contemporary, each of which is a sentimentalist theory of morality with empathy (or sympathy, in the case of the historical theories) at its heart. These are the sentimentalist moral theories of David Hume, Adam Smith and Michael Slote. I then apply the motivation objection to instrumental views of empathy’s role in ethics, before defending one particular instrumental view, according to which empathy can play a positive role in morality when integrated with virtue, and the virtue of compassion in particular.
Declaration

I hereby swear that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

The author of this thesis owns certain copyright or related rights in it (the “Copyright”) and she has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/)
intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on presentation of Theses.
1.0 Dedication and Acknowledgements

This thesis is dedicated to my mother, Barbara Impey, *sine qua non*.

I would like to thank my father, Martin Impey, who died suddenly in 2008 while I was writing this thesis. I know he would have been proud of me. I would also like to thank my brother, Alex Impey, and my step-mother, Pat Impey, for their support and encouragement.

I am hugely indebted to all my friends, but I would like to thank Paul Knights, Michael Rush, and Rebecca Reilly-Cooper in particular, for innumerable coffees and countless conversations, sometimes philosophical, always mirthful. I am also grateful to Katerina Bantinaki for suggesting that I do a PhD at the University of Manchester, and to Graham Stevens, who provided invaluable guidance on many occasions.

The thesis would not have been possible without the generous financial award, in the form of the Samuel Hall Studentship for Philosophy, which I received from the Philosophy Discipline Area at the University of Manchester during my second and third years of study.

Special thanks must go to my supervisors, Professor Peter Goldie and Dr. Joel Smith. It is an honour to have been supervised by Peter Goldie. Before his untimely death in October 2011, I benefited immeasurably from, among many other things, his kindness, his wisdom, and his firm belief in me. He continues to be an inspiration.
I owe my deepest gratitude to Joel Smith, without whom I would have been unable to complete this thesis. Joel provided exceptionally patient and diligent support, and was incredibly generous with his time. I benefited enormously from his trenchant, honest appraisals of my work, from his knowledge, his insight and his sense of humour. I cannot thank him enough.
1 Introduction

1.1 Thesis Overview

Research into empathy has blossomed in recent years. Empathy research continues to flourish in social and developmental psychology, moral philosophy, aesthetics, philosophy of mind, social and cognitive neuroscience, psychotherapy, education, and more besides. There is, however, much disagreement as to what exactly ‘empathy’ is, and much debate surrounding the exact role it plays in human life. In addition, many phenomena have appeared under the term ‘empathy’. Indeed, as social neuroscientists Tania Singer and Claus Lamm (2009, p. 82) have pointed out, ‘there are almost as many definitions of empathy as there are researchers in the field.’ Social psychologist Daniel Batson (2009a, p. 3) agrees: ‘the term empathy is currently applied to more than a half-dozen phenomena. These phenomena are related to one another, but they are not elements, aspects, facets, or components of a single thing that is empathy, as one might say that an attitude has cognitive, affective, and behavioural components. Rather, each is a conceptually distinct, stand-alone psychological state. Further, each of these states has been called by names other than empathy.\’

This thesis focuses on empathy’s role in ethics, where it is often considered to be of significant moral value. It can be morally valuable in two ways. Firstly, there are those who consider it to be somehow constitutive of morality, as forming morality’s very basis. The sentimentalist moral philosopher Michael Slote (2007; 2010a), for example, has argued recently that empathy is the sine qua non of morality. It is constitutive of morality, he claims, because it is necessary for caring about the well-being of others, and for making moral judgements. Slote claims that his contemporary view is a direct descendant of David Hume’s sympathy-based sentimentalist moral
theory. Hume also believed that sympathy, which is similar to what some call empathy today, is foundational for morality. And Hume’s fellow sentimentalist Adam Smith shared this view. For Smith, sympathy—also considered by some to be akin to empathy—operated differently but still played an essentially constitutive role in morality.

Secondly, there are those for whom empathy is of instrumental value for morality. Even if one does not consider it to be foundational for, or in some way constitutive of, morality, one might still argue that it is of instrumental moral value. There are various ways in which we can apportion instrumental value to empathy. It can be instrumentally valuable, for example, insofar as it can help us to understand another’s point of view. Cambridge professor of psychology and psychiatry Simon Baron-Cohen (2011) has recently gone so far as to claim that, such is empathy’s instrumental value, we must replace the term ‘evil’ with the term ‘lack of empathy’. When human beings are evil and cruel, he argues, it is because they lack empathy. Empathizing can prevent evildoing and promote peace. It is ‘the most valuable resource in our world’, a ‘universal solvent’ that can ameliorate, and even resolve, conflicts and disputes. (Ibid., pp. 103-104) Baron-Cohen is not alone in making such claims. Empathy is seen as instrumentally valuable by many of the world’s religions, as well as secular humanism, where its instrumental value lies primarily in the fact that it can guide moral action.

In this thesis I argue that empathy’s advocates, including both those who claim a constitutive relation between empathy and ethics, and those for whom empathy has instrumental moral value, face what I call the motivation objection. This states that those who make claims about empathy’s moral value fail to take into account the fact that empathy can be affected and
influenced by our motivations, taking this term in a broad sense. The fact is that empathy, as I will understand the term, is neutral when it comes to motive: in other words, we might empathize for morally good motives, and for morally dubious motives, or out of some other kind of motive, such as self-interest, which may or may not coincide with what is morally appropriate. Moreover, the emotions or behaviour that can result from the empathizing can be affected by the motive behind it, with the result that we can have an inappropriate emotional response or act inappropriately as a result of empathy. Considerations of motive are not made often enough by those who consider empathy to be of moral import, and this has consequences for the claims that are made about its ethical value.

I am using the term ‘motivation’—or ‘motive’, as I will also sometimes refer to it—in a broad sense in this thesis. I include under its rubric a person’s dispositions and character, as well as his beliefs, desires and intentions. A person’s motive for acting is, roughly, the reason why the person does the act. I will be assuming throughout the thesis—although I recognise that it is controversial to do so—that we commonly make moral evaluations of a person in terms of (i) his or her actions in themselves, (ii) the motives behind the actions, (iii) the person’s character, and (iv) the consequences of the actions.\(^1\) Empathy, as I will understand the term, is an act or process.\(^2\) If a person engages in an act, deliberately or spontaneously, we can say of the act itself that it was good or bad. And we can say the same about a person’s motive(s) for acting: we can call his motive good or bad, and for a variety of reasons. The evaluation we make of the act itself is influenced by our evaluation of the motive behind the act, although this need not be the case.

\(^1\) In taking this approach, which I believe to conform to our everyday moral thinking, I appreciate that I am registering a disagreement with many moral theories, such as consequentialism, which holds that only consequences matter, and with Kantianism, which holds that only motives matter.

\(^2\) I will be saying more about empathy as a process in §1.2.3.
And a person’s motive, if it manifests itself in his behaviour often enough, can shape our evaluations of that person’s character as, for example, virtuous or vicious. We can also evaluate an act in terms of its consequences.

I argue that our moral evaluations of empathy can be made in all these ways: by considering the act or process of empathizing; the motives behind the empathizing; the character of the empathizer, and the consequences of the empathizing. I argue that, because empathy can be affected and influenced by our motivations, we ought to take this fact into account when considering its moral value. The motivation objection claims that holders of both constitutive and instrumental views of empathy’s moral value fail to do so.

There are two main parts to the thesis. Part I reflects constitutive views of empathy’s moral value, and Part II reflects instrumental views. In Part I, which comprises Chapters 2, 3 and 4, I evaluate three sentimentalist moral theories—two historical and one contemporary—all of which are versions of the view that there is a constitutive relation between empathy and ethics. They are the sentimentalist moral theories of Hume, Adam Smith and Michael Slote. These three theories all place empathy (although Hume and Smith employ the term ‘sympathy’\(^3\)) at their heart. Hume, Smith and Slote each consider empathy to be constitutive of morality because, for example, empathizing is sufficient for feeling compassion, and is necessary for moral judgement. Furthermore, it has been argued that Hume and Smith were early simulation theorists. Their notions of sympathy, it has been claimed, are methods of attributing mental states to other people. In addition, Smith and Slote attempt to use empathy to define aspects of morality in a non-

---

\(^3\) Hume and Smith speak of sympathy and Slote of empathy. All three concepts differ in various respects, but share important similarities. In this thesis I will be using the term ‘empathy’ rather than ‘sympathy’, unless I refer to Hume’s and Smith’s notions specifically.
circular way. For example, Slote attempts to define virtue in terms of empathic concern, which, as we will see in §1.2 below, is a product of empathy processes. I argue that each sentimentalist faces the motivation objection. The motivation objection has the following three dimensions or aspects.

(1) **Reliability:** empathy processes are not reliable insofar as they do not always lead to the right result(s). This aspect of the motivation objection applies in particular to claims that are made about empathy and its role in third-person emotion attribution. I argue that a person’s motives, including his disposition, character, desires and beliefs, mean that empathy processes cannot be relied upon to produce correct third-person mental state attributions.

(2) **Function:** empathy is not sufficient for various functions. For example, empathy is thought to be sufficient for causing compassion: if we empathize with someone who is suffering, it is argued, we will feel compassion for them. Another alleged (constitutive) function of empathy is enabling moral judgement: if we empathize, we will feel moral sentiments of approval or disapproval. However, I argue that empathy is, in fact, insufficient for fulfilling these—and several other—functions because of motive: there are many ways in which our motivations can influence the outcomes of empathy processes.

(3) **Circularity:** empathy is neutral about motive so using it to define or explain certain aspects of morality in a non-circular way is problematic. Smith and Slote face this aspect of the motivation objection because each seeks to define some aspect of morality using empathy (or sympathy, in
Smith’s case). Smith claims that what it is to be virtuous can be explained in terms of sympathy, and Slote attempts to explain moral approval and disapproval in terms of empathy.

As well as applying the motivation objection to the sentimentalist theories, I also demonstrate how some of the interpersonal phenomena that we call empathy today are thought to constitute morality. Although Hume and Smith refer to sympathy and not empathy, it is frequently claimed that they were early observers of interpersonal phenomena that are sometimes now called empathy phenomena, and examples from their work have been utilised by philosophers, psychologists and neuroscientists, to name but a few, to buttress assertions about the workings of various interpersonal phenomena. In the philosophy of mind, for example, Alvin Goldman (2009c, p. 433) has argued that Hume and Smith ‘offered subtle and prescient insights into these [...] phenomena, but only in recent years have contemporary philosophers and scientists of the mind begun to rediscover, substantiate, and refine those insights.’

Hume and Smith could not have been aware of the phenomena people have claimed to observe in their work, but they are nevertheless recruited to reinforce or illustrate claims about empathy in many research domains. This makes looking at their notions of sympathy highly relevant to today’s debates on empathy. Slote allies his sentimentalism more with Hume’s than Smith’s. Because he bases his notion of empathy on contemporary research undertaken in social and developmental psychology, in particular the work of Martin Hoffman (2000) and Daniel Batson (1991), considering Slote’s theory will enable us to see not only the historical parallels between the earlier sentimentalist views and his own, but also the interdisciplinary nature of the empathy concept, and some of the common assumptions about
empathy that are made by the three sentimentalists. Hume, Smith and Slote use broad concepts of empathy: they tend to run together what are in fact distinct phenomena under this rubric. Using broad empathy concepts is no bad thing in itself. But sometimes it can lead to claims being made about what empathy is capable of doing that are unwarranted.

Part II of the thesis comprises Chapters 5 and 6. In Chapter 5, I target various versions of the view that empathy has an instrumentally valuable role to play in ethics, and apply the motivation objection to them. Empathy is often considered instrumentally valuable for morality in terms of its output, for example feelings of sympathy or compassion for the suffering of others, and a motivation to alleviate that suffering. However, there are other ways in which it connects to motive that have received little attention in the literature. Empathy, for example, can itself be motivated, and not always for the good, which can affect the outcome of the empathy, for example the feeling or the helping behaviour, and how we evaluate these. In Chapter 5 I use examples and current empirical evidence, where this is relevant, to apply the motivation objection: I argue that consideration of motive is vital when making claims for empathy’s instrumental moral value.

Taken together, Chapters 2, 3, 4 and 5 constitute significant evidence against both constitutive and instrumental views of empathy’s moral value. However, they should not be taken to imply that empathy can play no positive role in morality. On the contrary: the final chapter, Chapter 6, defends a particular interpretation of an instrumental view. I argue that, when properly integrated with virtue, in particular the virtue of compassion, empathy can play an important, instrumentally valuable, role in our moral lives. Empathy can serve to enhance compassion by ensuring, for example, that we are more likely to feel the right emotional response to another’s
suffering. It can ensure that the virtuous person is more likely to help another person who is in need, and that he or she is more likely to help in the right kind of way—that is, in the way that is appropriate to the virtue of compassion—by taking the feelings of the other person into consideration. Chapter 6 also serves to justify the circularity aspect of the motivation objection, for it argues that virtue must precede empathy, in order for empathizing to be for the good.

With this synopsis of the thesis in mind, I now provide a brief overview of the empathy concept.\textsuperscript{4} There is a wealth of empathy research, and the concept has a rich history in many discipline areas. I do not have the space here for a review of empathy as it appears in all the research domains. I therefore include areas in which empathy research is and has been undertaken that are relevant for my thesis, and introduce some of the phenomena that have been called empathy, to which reference is made in the thesis. I introduce what I will be referring to—following Alvin Goldman (2006)—as ‘low-level empathy’ (LLE) and ‘high-level empathy’ (HLE), clarify these terms, and justify dividing empathy phenomena in this way rather than in some other way, which it is possible to do. For instance, we could divide empathy phenomena according to whether they are automatic or controlled, or whether they are conscious or unconscious, or whether they are what psychologists refer to as ‘cognitive’ or ‘affective’ in nature. I do not offer a definitive description of empathy: defining empathy is notoriously difficult, and there are many definitions of empathy and related phenomena in the literature. Rather, I draw a distinction between HLE and LLE phenomena, which enables me to refer to phenomena which, broadly speaking, fall under each of these categories, without adding to the myriad

\textsuperscript{4} See the introductory chapter of Coplan & Goldie (2011) for a comprehensive overview of the empathy concept and its history.
empathy definitions already extant in the literature. I ‘define’ empathy phenomena as high- or low-level by distinguishing them, roughly speaking, according to whether or not the phenomena involve imaginative perspective-taking.

1.2 Empathy Phenomena

Empathy research is undertaken in many areas and is often inter-disciplinary in nature. There are many definitions of empathy, and many phenomena fall under this general rubric. In addition, some empathy concepts are broad, including distinct phenomena under the umbrella term ‘empathy’ (e.g. Hoffman, 2000; Preston & de Waal, 2002) whereas others refer to just one phenomenon as empathy (e.g. Goldie, 2000). Although many of the phenomena that we now call empathy can be traced back at least to the work of Hume and Smith in the eighteenth century, the word ‘empathy’ is itself relatively new. It was not until 1909 that Theodore Lipps’ (1903) notion of *einfühlung* was translated by Edmund Titchener (1909) to give us the English word ‘empathy’. *Einfühlung* literally means ‘feeling into’, and for Lipps constituted a process of ‘inner imitation’, a way of experiencing a work of art or some other object in the external world by mentally projecting oneself and one’s feelings onto the object, striving—as Lipps put it—to perceive the object as it is in itself, to unconsciously simulate it using our own minds and bodies. Lipps developed his theory of *einfühlung* further to explain how it is that we come to understand the mental states of other people, taking inspiration from Hume’s notion of sympathy. After having translated the *Treatise* into German, Lipps took from Hume the idea that some process—sympathy, in Hume’s case—enables men’s minds to somehow ‘mirror’ each other. (Montag, et al., 2008; *Treatise*, p. 365) To illustrate *einfühlung*, Lipps used the example—also employed by Smith to illustrate his notion of

5 See Wispé (1987) for a detailed exegesis of the concept of *einfühlung* and its historical roots.
sympathy—of watching an acrobat walking on a tightrope. The observer inwardly imitates what the acrobat is thinking and feeling, and does so unconsciously.

Considerably before the appearance of the word empathy in the twentieth century, and the publication of Lipps’ work on *einfühlung*, the notion of sympathy was employed by Hume and Smith in the eighteenth century. Hume used the term to explain, for example, how we respond emotionally to the suffering of others, how we come to have moral and aesthetic responses, and how our emotions and opinions can come to be transmitted between us. Briefly, Humean sympathy is a psychological mechanism that converts an observer’s idea of a target’s emotion into that emotion in the observer.⁶ Because it appears, on the face of it, to be an automatic process that involves the instantaneous transmission of sentiments from one person to another, Hume’s sympathy has been seen as an early version of what we now refer to as *emotional contagion*. Emotional contagion has been defined as ‘the tendency to automatically mimic and synchronize facial expressions, vocalizations, postures, and movements with those of another person and, consequently, to converge emotionally’. (Hatfield, et al., 2009, p. 19) Contagion usually results in coming to feel as another person feels. (Batson, 2009, p. 5)

*Perspective-taking* forms of empathy are often traced back to Smith. The following example, from Smith’s *Theory of Moral Sentiments*, mentioned above during the discussion of *einfühlung*, is often cited as an early example of the low-level empathy phenomenon of motor mimicry, about which I will say more below.⁷ However, on Smith’s view, sympathy is a process of

---

⁶ I provide a detailed exegesis of Hume’s notion of sympathy and its workings in Chapter 2.
⁷ See, for example, Goldman (2006, p. 17), Batson (2009, p. 6) and Hoffman (2000, p. 37).
imaginative perspective-taking, invoked to explain how we respond emotionally to the suffering of others and how we make our moral evaluations of their actions and characters.\(^8\)

The mob, when they are gazing at a dancer on the slack rope, naturally writhe and twist and balance their own bodies, as they see him do, and as they feel that they themselves must do if in his situation. (TMS, p. 10, I.i.1.3)

On Smith’s view, sympathy usually refers to the imaginative process of putting ourselves in another person’s shoes. Smith also speaks of another form of perspective-taking, however, which involves taking on the perspective of another person in the imagination. Both forms of perspective-taking have been referred to as types of mental simulation by philosophers of mind (see e.g. Goldman, 2006; Gordon, 1995), and both are forms of high-level empathy, which is the focus of §1.3 below.

Closely related to contagion, and sometimes thought even to underpin it, is the phenomenon of motor mimicry, sometimes also known as imitation. As we will see below, some refer to these as ‘mirroring’ processes, because they are apparently subserved by so-called ‘mirror neurons’.\(^9\) Humean sympathy is sometimes seen as an early form of mimicry. (See e.g. Goldman, 2006, p. 17) Imitation or mimicry can be described as ‘adopting the posture or matching the neural responses of an observed other.’ (Batson, 2009, p. 4)

According to Hoffman (2000, p. 37), mimicry is one way in which a person’s empathic reactions can be aroused when confronted by the distress of another person. On his view, mimicry is a combination of two steps: imitation and feedback. When an observer imitates a target, he ‘first automatically imitates and synchronizes changes in his facial expression,

\(^8\) I provide a detailed exegesis of Smith’s notion of sympathy in Chapter 3.

\(^9\) See Decety & Meltzoff (2011) for an introduction to the relationship between empathy, imitation and mirror neurons.
voice and posture with the slightest changes in another person’s facial, vocal and postural expressions of feeling.’ (Ibid.) These changes trigger an afferent feedback response in the observer that produces feelings in him that match those being experienced by the target. The physiological changes the observer undergoes by imitating the target’s face, voice or posture cause that observer to feel the emotion the target is feeling. This process happens innately, involuntarily and unconsciously. (Ibid.)

The study of empathy phenomena has been greatly enhanced by advances in the fields of cognitive-affective and social neuroscience, where research centres primarily on the neural processes and mechanisms underlying these phenomena. Over the past twenty years, neuroscience research has provided a wealth of empirical data that has been pivotal in enhancing our understanding of the nature and role of such phenomena, and the neurophysiological processes and mechanisms subserving them. (Cf. Coplan & Goldie, 2011, p. XXVIII) Research has shed light, for example, on the neurological processes underlying so-called empathy deficits in psychopaths and people with autism, on what happens in our bodies when we observe others in pain, and on the way empathy phenomena contribute to human interaction and understanding. Hugely important in this latter regard has been the discovery of, and subsequent research into, so-called mirror neurons.

Mirror neurons were discovered in the ventral premotor cortex (brain area F5) of macaque monkeys by researchers at the University of Parma in Italy in 1992, although it was not until 1996 that the term ‘mirror neurons’ was adopted to describe these new cells. (Iacoboni, 2011)10 Mirror neurons are

10 See Gallese, et al. (1996) for the original research findings.
cells that fire in a monkey’s brain both when it is undertaking goal-directed actions like grasping a cup or peeling a banana, and when it is observing a conspecific (or human) doing those same actions. The observing monkey’s neurons fire as though the monkey were itself undertaking the action it is observing: its own neural cells ‘mirror’ or simulate those of the other. The existence of a mirror neuron system (MNS) in humans was hypothesised in the early research on mirror neurons in monkeys. (Gallese, et al., 1996) However, it is difficult to do ‘single-cell’ recordings on humans, as it is a highly invasive procedure, so evidence for a human MNS depends largely on the use of neuroimaging techniques such as functional magnetic resonance imaging (fMRI) and transcranial magnetic stimulation (TMS), which means research into the human MNS is arguably less advanced and less well-established. (Pfeifer & Dapretto, 2009) Nonetheless, it is now widely believed that there is a mirror neuron system in humans, and, moreover, that there is a connection between the human MNS and empathy. (Ibid., p. 187) I will say more about this connection in §1.3. Neuroscientific research is not limited to the study of the neural mechanisms behind phenomena like contagion, mimicry and imitation. Perspective-taking forms of empathy are also studied by neuroscientists, although there is little evidence of perspective-taking being subserved by mirror systems.11

Empathy phenomena continue to be the subject of empirical research in social and developmental psychology. Psychologists usually consider empathy to involve some sort of affective response to the suffering or perceived need of another person, although they disagree as to whether empathizing requires feeling something that is analogous to what the other person is feeling, or whether it is sufficient for empathy to feel an emotional response that is more ‘congruent’ with another person’s situation than one’s

11 But cf. Goldman (2011) for a different view.
own. In other words, there is disagreement over the extent to which empathizing involves the sharing of affect. One influential theory of empathy, put forward by Hoffman (2000), holds that human empathy is a set or cluster of capacities that develops along a timeline from birth, from motor mimicry and imitation in neonates and babies to perspective-taking in adolescents and adults. These capacities, Hoffman claims, enable us to feel other-oriented emotions when we observe suffering in others, and can motivate us to act to alleviate that suffering. The ‘other-oriented’ emotions all stem from what he calls ‘empathic distress’, and manifest themselves as, for example, sympathetic distress (akin to sympathy), compassion, and empathic anger (anger on behalf of another person). (Ibid.)

Social psychologist Daniel Batson’s (1991) well-known research on empathy posits a direct link between empathy and prosocial behaviour. His ‘empathy-altruism hypothesis’ holds that both low- and high-level empathy phenomena (like perspective-taking) can lead directly to altruistic helping behaviour. ‘Empathy’ in the ‘empathy-altruism’ hypothesis is what Batson and others call empathic concern, which is an other-oriented (i.e. felt for the other person) attitude or feeling of concern for the well-being of the person in need that often results in altruistic helping behaviour. This is distinguished by psychologists from helping behaviour that results from personal distress. Psychologist Nancy Eisenberg (2009, p. 72) defines personal distress as a ‘self-focused, aversive affective reaction to the apprehension of another’s emotion, associated with the desire to alleviate one’s own, but not the other’s distress.’ Helping that results from personal distress is not altruistic. Rather, it is egoistic: we help the victim in order to alleviate our own distress rather than the victim’s. It is important to note the purported connection between emotions like sympathy and compassion and

---

12 See, for example, Eisenberg & Strayer (1986) and Hoffman (2000).
empathic concern. Psychologists often consider some of the feelings or states that are induced by empathy phenomena like perspective-taking or mimicry—empathic concern, for example—to be similar to sympathy or compassion insofar as they are all emotions felt for another person: they are other-oriented, rather than self-oriented.

Emotional contagion, imitation and mimicry are all forms of what I will be referring to in the thesis as low-level empathy (LLE). The two types of perspective-taking I mentioned are both forms of what I will be referring to as high-level empathy (HLE). I now explain and justify making this distinction between LLE and HLE and, in so doing, provide an introduction to empathy and the mental simulation debate in the philosophy of mind. I also argue that empathy is a process, not an outcome, that high-level empathy is not a virtue, and that it can be both spontaneous and deliberate.

1.3 High- and Low-level Empathy

I distinguish between the phenomena that have been called empathy according to whether they involve the imagination or not. It is not uncommon to make such a distinction: phenomena that have been called empathy, such as emotional contagion, mimicry, and imitation, are frequently distinguished from types of empathy that involve imaginative perspective-taking. This distinction is, however, expressed in different ways, in different disciplines. In psychology, for example, the distinction is often marked by referring on the one hand to ‘affective’ empathy, and on the other to those types of empathy that are more ‘cognitive’, which is to say that they employ ‘higher-order’ cognitive faculties like imagination and memory. (See e.g. Baron-Cohen, 2011; Hoffman, 2000) However, this distinction is not specific enough for my purposes, because the term ‘cognitive empathy’ can have multiple referents. That is, it can be used to refer to processes other
than perspective-taking. (Batson, 2009, p. 4) Given that empathy phenomena can be reasonably grouped together according to the relative involvement of the imagination, I adopt Goldman’s distinction between high- and low-level empathy, as he makes his distinction along these lines. Before introducing empathy and simulation theory, I will argue that empathy, whether high- or low-level, is best understood as a process, and not an outcome.

1.3.1 Process vs. Outcome

Empathy is sometimes referred to as a product or outcome. This is often an emotion—feeling what the other is feeling, a token of the same type of emotion that they are feeling. Jesse Prinz (2011, p. 212), for example, thinks empathy is an outcome or product—‘feeling what one takes another person to be feeling’—which can be generated by either of two processes: either ‘automatic contagion or [as] the result of a complicated exercise of the imagination.’ Singer and Lamm (2009, p. 82) also define empathy as an outcome: empathy is ‘an affective response to the directly perceived, imagined, or inferred feeling state of another being.’ And Julien Deonna (2007) claims that empathizing involves knowing how the other feels and feeling something similar to what they are feeling.

Empathy is also described as both a process and an outcome. On Hoffman’s view, for example, empathy is ‘a vicarious emotional response to another person’. (Hoffman, 2000, p. 30) But Hoffman actually defines empathy ‘not in terms of outcome (affect match) but in terms of the processes underlying the relationship between the observer’s and the model’s feeling.’ (Ibid., p. 30) So what determines whether or not we have empathized is whether we have undergone a particular psychological process and that this has resulted in a vicarious emotional response. This is supposed to rule out someone’s just happening to be experiencing the same emotion as the target coincidentally, rather than as the result of having empathized with them. Stephen Darwall
(1998, p. 261) also refers to empathy as both a process and an outcome: ‘empathy consists in feeling what one imagines he feels, or perhaps should feel (fear, say), or in some imagined copy of these feelings, whether one comes thereby to be concerned for the [other person] or not.’

I refer to empathy as a process because calling the output ‘empathy’ implies that there is just one output state, empathy, when in actual fact the outcomes—the ‘empathic’ states—can be very diverse, especially in the case of high-level empathy. An empathic state can be one that matches the target’s state, but this can be to a greater or a lesser degree. We can fail to feel anything as a result of empathy, and frequently come to feel something totally different to the target. For example, both perspective-taking and low-level empathy processes can cause feelings of personal distress as well as empathic concern. So it does not seem to make sense to call the outcomes of the empathy processes empathy, because this implies that they are all copies of the target’s state, when this need not be the case.

1.3.2 Empathy and Simulation Theory

In the philosophy of mind, the term empathy is used to describe third-person emotion state attribution, also known as emotion mindreading, which involves some kind of simulation process. The mental simulation debate in the philosophy of mind concerns how it is that we are able to make mental state attributions to other people. This so-called ‘mindreading’ ability was, until the emergence of the simulationist challenge in the mid-eighties, largely explained by appeal to the possession and deployment of a (tacit) psychological theory, a stock of conceptual knowledge and generalizations about other people that we use to reason about others’ minds. The idea of mental simulation was put forward—indeently—by

Simulation theorists hold that we mindread by internally simulating—in the sense of duplicating or replicating—other people’s mental states. (Goldman, 2006, p. 36) Instead of using a theory, an internal body of information that we could use to attribute mental states to others, and using this information to predict what they might do, we simulate or internally replicate what others might do in a given situation. For example, A might predict a decision of B’s by putting himself in B’s shoes. A pretends to have the same initial states as B, for example pretending to believe something B believes, or to desire something B desires. A then makes a decision using his own decision-making mechanisms—he makes a decision based on the pretend beliefs while he is still pretending. A then attributes that (pretend) decision to B. (Cf. Goldman, 2006, p. 19)

Empathy, when utilised in relation to mindreading, concerns the third-person attribution of emotion states: that is, the attribution of emotions to other people. The simulation processes hypothesised to underlie emotion mindreading can be distinguished using different terminology, and according to how the processes are conceptualised. On Goldman’s view (2006, p. 207), emotion mindreading is best explained by his ‘dual-process’ theory of empathy, which reflects his duplex theory of simulation-based mindreading. According to Goldman, there are two ‘routes’ to emotion mindreading, a higher-level route and a lower-level route, each of which involves a different type of simulation. On Goldman’s view, empathy

---

¹³ See the collection by Davies and Stone (1995) for an introduction to the mental simulation debate, and these early papers by Heal, Goldman and Gordon. Heal favours a simulation process that she calls ‘co-cognition’, which she distinguishes from Goldman’s and Gordon’s conceptions of simulation. See also Heal (2003), especially Chapters 2 and 6.
describes the outcome of these two simulation processes. The lower-level route or process involves mirroring, which, according to Goldman, consists in ‘unmediated resonance’ with another person’s emotions. The higher-level route involves so-called ‘enactment imagination’, abbreviated by Goldman to ‘E-imagination’. (Goldman, 2006, pp. 127, 149)

Unmediated resonance or mirroring is Goldman’s favoured low-level simulation process. He calls it mirroring because, he claims, it involves recruitment of the MNS. It operates in emotion mindreading as follows. When someone is experiencing a basic emotion—disgust, for example—then the ‘perception of the target’s face “directly” triggers (subthreshold) activation of the same neural substrate of the emotion in question.’ (Ibid., p. 127) This leads to the empathizer feeling disgust, and then attributing that emotion to the target. High-level empathy, by contrast, involves imagining oneself into the target’s shoes: in other words, taking on his perspective in the imagination. Goldman claims that high-level empathy is subserved by E-imagination, a process that involves enacting the mental state of another person in a simulation, rather than supposing that you are in the same mental state as them. As Goldman puts it, ‘When I imagine feeling elated, I do not merely suppose that I am elated; rather, I enact, or try to enact, elation itself.’ (Ibid., p. 47, Goldman’s italics)

Other simulation theorists offer slightly different accounts of the processes underlying emotion mindreading and use different terminology to describe them. Karsten Stueber (2006, p. 170), for example, refers to ‘basic’ empathy rather than lower-level empathy, and to ‘reenactive’ empathy as opposed to high-level empathy. Basic empathy, according to Stueber, involves ‘underlying psychological mechanisms’ (ibid., p. 170) and reenactive empathy involves ‘using our own cognitive and deliberative capacities in
order to re-enact or imitate in our own mind the thought processes of the other person.’ *(Ibid., p. 21)* Frédérique de Vignemont (2009, p. 460) refers to ‘mirroring’ empathy and ‘reconstructive’ empathy. Mirroring empathy is, on her view, ‘induced by the observation of emotional cues that induce emotion mirroring’; reconstructive empathy ‘is induced by the pretense of the emotional situation.’ *(Ibid.)* Robert Gordon (1995) distinguishes between ‘facial empathy’—which includes mimicry and contagion—and ‘higher-forms’ of empathy, which include imagining being the target of the simulation.14

Although each of the definitions above employs different terminology to describe the processes underlying emotion mindreading, these processes seem to fall naturally into two camps, which appear to share common features. For example, low-level empathy, facial empathy, mirroring and basic empathy require no input from the imagination. These phenomena seem, broadly-speaking, to consist in a kind of ‘resonating’ with another’s emotions, of coming to feel what the other person is feeling by ‘picking up’ that person’s emotions without being consciously aware of what caused this to occur. On the other hand, reconstructive, reenactive and high-level empathy all involve the imagination, seem to be more effortful, and may involve more advanced cognitive processes, like memory, than do their lower-level counterparts. Given that all the definitions of the processes claimed to underlie emotion mindreading have in common that they either do or do not involve the imagination, I follow Goldman by using the terms low- and high-level empathy. Goldman, however, did not always make his distinction in this way. In *Simulating Minds* (2006), he offered criteria for

---

14 I do not want to adjudicate here as to who is right about how to conceptualise the processes believed to underlie simulational emotion mindreading. Nor do I intend to comment on the extent to which simulation is involved in emotion mindreading, or to claim that it is the default way in which we mindread others’ emotions. I am simply assuming here that simulation processes, howsoever they are conceived, play some role in emotion mindreading.
distinguishing between low- and high-level simulation that were challenged by de Vignemont, prompting Goldman to revise his early definition. A brief examination of the definition, and de Vignemont’s objections, will be instructive because it will introduce an idea that will be of great importance later on in the thesis, viz. the fact that both high- and low-level empathy can be modulated.

On Goldman’s early definition (2006, p. 113), simulation processes are low-level if they are ‘comparatively simple, primitive, automatic, and largely below the level of consciousness.’ They are simple because they target ‘basic’ mental states, as opposed to ‘mental states of a relatively complex nature, such as propositional attitudes’ like beliefs and desires. (Ibid., p. 147) They are automatic insofar as they are not subject to voluntary control, the implication of this being that they cannot be modulated or interfered with by any knowledge that the simulator possesses. (Ibid., p. 21) They are unconscious because we are not aware of ourselves undergoing them. Simulation processes that are higher-level, on the other hand, are ‘potentially and intermittently under intentional guidance and control’ (Ibid., p. 132) and target complex mental states like propositional attitudes. They also have ‘some degree of accessibility to consciousness.’ (Ibid., p. 147)

De Vignemont (2009, p. 460) argues, however, that low-level empathy cannot be a purely automatic process because it can actually be mediated, regulated, and guided by knowledge that the empathizer possesses. It cannot be defined as ‘unmediated resonance’ as Goldman claims because it can, in fact, be mediated. For example, Singer, et al. (2006) showed that our relationship with the target of the empathy modulates our low-level empathy by influencing our affective response to his or her pain: if you believe that a person who is in pain treated you unfairly, you are less likely to low-level
empathize to the same extent with him or her. In addition, our responses to the pain of others can be controlled, as is the case with medical practitioners who inhibit their LLE when they observe their patients experiencing pain. (Cheng, et al., 2007) And if someone is in pain and we know that the pain is necessary, to cure an illness, for example, we feel weaker responses as a result of low-level empathy than we do when we believe the pain being experienced to be unnecessary. (Lamm, et al., 2007)

De Vignemont also disputed Goldman’s early distinction between high- and low-level empathy according to their supposed voluntary/involuntary nature. Goldman originally claimed that LLE cannot be voluntarily triggered. De Vignemont agrees that we cannot decide to activate our MNS when we perceive another person undergoing an emotion: it just happens to us. However, she rightly points out that there is a sense in which we might want to call HLE involuntary, too. This is because, although it is true that we can decide to engage in HLE, we cannot always help but feel an emotion when we do so. So in this sense, HLE is not voluntary.

When I feel sad with someone, it is not because I want to be sad with her, it just happens to me. And as much as I may wish to share someone’s happiness, it does not necessarily work. There is more to reconstructive empathy than voluntary perspective-taking. (Vignemont, 2009, p. 461)

In light of these criticisms, Goldman redefined low-level and high-level simulation processes as follows, dropping the automatic/controlled and voluntary/involuntary distinctions.

A low-level simulational event is a mental (or neural) event produced by the subject’s observation of a suitable behavioural manifestation of a matching event (or a behavioural manifestation that typically results from a matching event). A high-
level simulational event is an event produced by the imagination (usually in concert with one or more other subsystems). (Goldman, 2009b, p. 485)

The “level” classification of any individual case [of simulation] depends on what initiates activation in the given brain areas. If it is initiated by observation of another actor, it is low-level simulation. If it is initiated purely by imagination, it is high-level simulation. If — in rare cases — an activation is overdetermined, and produced by both, then it is both low-level and high-level. (Goldman, 2009b, p. 485)

However, as Goldman (2009b, p. 485) acknowledges, his revised definition presupposes his own conceptions of simulation processes, *viz.* mirroring and E-imagination. I do not want to adjudicate between de Vignemont and Goldman as to who is right about the mechanisms underlying low-level empathy: this must be determined by the empirical evidence.¹⁵ I will include mirroring as one possible low-level empathy phenomenon. It is sufficient for my purposes that a plausible way to distinguish between empathy phenomena is according to whether they involve the imagination or not.

1.3.3 High-level Empathy

1.3.3.1 Types of High-level Empathy

I incorporate two forms of perspective-taking in my loose definition of HLE. These are what I am calling *in-his-shoes imagining*—following Peter Goldie (2000)—and *imagining being the other person*.¹⁶ They reflect Goldman’s favoured form of simulation and also the type of simulation favoured by Gordon (1986; 1995). Both constitute forms of high-level empathy on my definition because both are caused by the imagination. The incorporation of both types of process as types of empathy is widespread in the psychological literature: Hoffman (2000) makes a similar distinction and includes both

---

¹⁵ See Jacob (2008) for a review of the contribution mirror neurons make to emotion mindreading and an objection to Goldman’s claim that there is a link between mirroring and mindreading. See Goldman (2009a) for Goldman’s reply.

¹⁶ Also see Williams (1973), Velleman (1996) and Wollheim (1973) for more on the distinction between these two forms of imagining.
types of imagining in his broad conception of empathy, and Batson, et al. (1997) distinguish between imagine-other and imagine-self perspective-taking, which are similar to the in-his-shoes imagining and imagining being the other person that I have in mind.

Goldie (2000) separates the two phenomena into in-his-shoes imagining and empathy. On his view, ‘empathy is a process or procedure by which a person centrally imagines the narrative (the thoughts, feelings, and emotions) of another person,’ (ibid., p. 195) whereas in-his-shoes imagining is putting yourself in another person’s shoes, rather than imagining being that person. When I talk about imagining being the other person, it is something like Goldie’s conception of empathy that I have in mind. When I talk about in-his-shoes imagining, I follow Goldie here as well. Both are forms of high-level empathy insofar as they are caused by the imagination.

There are imaginative processes directed towards others’ psychological states that do not fall under the high-level empathy umbrella, even though they involve the imagination. One example is what Goldie (2011, p. 306) has called ‘imagining-how-it-is’. When imagining-how-it-is, the imaginer does not imagine from another person’s perspective, whether that is her own or the target’s. Rather, he or she imagines what a certain situation will be like for a certain person (or thing) in general, and not from any personal perspective. It is the fact that in-his-shoes imagining and imagining being the other person are both forms of perspective-taking that makes them candidates for high-level empathy.
1.3.3.2 Spontaneous and Deliberate High-level Empathy

High-level empathy is usually seen as a conscious, effortful process, something a person embarks upon, an intentional mental act. But sometimes it is not a process that is intentionally engaged in; we can find ourselves doing it without thinking. This distinction between high-level empathizing that we find ourselves doing without thinking, and that which we deliberately undertake, is well captured by Kendall Walton’s (1990) notions of spontaneous and deliberate imagining.

Imagining seems, in some cases, more something that happens to us than something that we do. Like breathing, imagining can be either deliberate or spontaneous. (Walton, 1990, p. 14)

Spontaneous imagining is ‘likely to be a more “vivid” or “realistic” experience than deliberate imagining [...] which, in its independence of the will, is more like actually perceiving or otherwise interacting with the real world.’ (Walton, 1990, p. 14) When we imagine deliberately, we are aware of the imagined world as having been deliberately imagined. But spontaneous imaginings ‘have a life of their own’: the imaginer is a ‘spectator rather than a perpetrator’ of her imaginings. (Ibid.) Walton argues that spontaneous imaginings tend to be more vivid than deliberate imaginings because, when we imagine deliberately, it is difficult to ignore the fact that we are engaged in imagining. But if we imagine spontaneously, the fact that we are imagining is not ‘forced’ into our occurrent thinking. We are not forced to dwell on the fact that we are imagining a particular thing, even though we would surely acknowledge that we were engaged in imagining it if we were asked. So our spontaneous imaginings are likely to be more vivid because they pop unexpectedly into our heads, and we are not simultaneously preoccupied with doing the imagining itself.

17 Cf. e.g. Williams (1973, p. 32), Wollheim (1984, p. 73) and Goldman (2006, p. 149).
Walton’s discussion pertained to imagining generally, but just as we can generally imagine spontaneously or deliberately, so can we also HLE spontaneously and deliberately. Making a distinction between HLE that is deliberate and HLE that is spontaneous is intended to capture the fact that HLE is an act of the imagination that can sometimes be engaged in for explicit reasons, but can also be something that we can find ourselves doing, and that can result in our feeling something as a result of this process. This will be important throughout the thesis. Imagining is an action, but it is not always done for explicit reasons by the imaginer. Calling an act of imagination that is not done for explicit reasons involuntary, as opposed to spontaneous, is not satisfactory: low-level empathizing is involuntary in the sense that it is an event that one passively undergoes, like a knee-jerk. But it can never be done for an explicit reason by an agent in the same way that high-level empathizing can. Low-level empathy can never be deliberate, for it is not directly subject to the will, but high-level empathy can.

1.3.3.3 High-level Empathy and Virtue

High-level empathy is a morally neutral ability, not a moral virtue. It is sometimes considered to be a moral virtue, but it is not a suitable candidate for being one: unlike a virtue, HLE is a capacity that can be used for good or ill. A virtue is a character trait of an agent. Agents can have many qualities, ranging from a good singing voice to benevolence and open-mindedness. (Battaly, 2011, p. 288) Virtues like benevolence and open-mindedness are particular types of good quality, viz. good qualities of character. More specifically, they are ‘dispositions of appropriate action, emotion, perception, and motivation.’ (Ibid.) We can acquire virtues, and they are voluntary, which is to say that ‘we exert some control over their acquisition and operation.’ (Ibid.) Moral virtues, such as courage, benevolence and temperance, are ‘acquired states of character that lie in a mean.’ (Ibid.) According to Heather Battaly’s (ibid.) Aristotelian view, ‘the benevolent
person helps the appropriate people, at the appropriate times, in the appropriate ways, because of appropriate motivations.

Intellectual virtues include open-mindedness, intellectual courage and intellectual autonomy. The open-minded person ‘appropriately considers alternative ideas, and does so because she is appropriately motivated.’ (Ibid., p. 289) Simply acting benevolently or open-mindedly is insufficient for possessing the virtue. To be benevolent, an agent must perform a benevolent action ‘for its own sake, that is because she desires the good and (correctly) thinks that virtuous acts are good.’ (Ibid.)

HLE can be voluntary, but this does not mean it is a virtue. Firstly, as I will argue in Chapter 5, HLE can be used for good or ill, and this is not the case with virtue. A torturer can use his ability to HLE to ascertain how best to torture his victim. Secondly, unlike virtue, one can choose not to exercise the skill of HLE when it is appropriate without this detracting from one’s having the skill; whereas if a person chooses not to act benevolently when appropriate, this detracts from our willingness to attribute the virtue to him. Thirdly, an agent has to be motivated to act virtuously because he believes that so acting is good in itself, rather than because he has some ulterior motive. His virtuous action must be motivated in ways appropriate to the virtue. So, if someone acted benevolently, for example, in order to impress his girlfriend, rather than because he believes that acting benevolently is good in itself, then he does not possess the virtue. But a person can be motivated to HLE for all sorts of reasons, as I will argue in Chapter 5, not all of which are good, and yet this would not detract from his ability to HLE.18

---

18 There are many other differences between a skill and a virtue, but these are all I need to discuss here.
1.4 Summary

In this introductory chapter I provided an overview of the thesis and set out my main objection—the motivation objection—to constitutive and instrumental views of empathy’s role in morality. I introduced the empathy concept and described some phenomena that have been called empathy. I explained and justified dividing empathy phenomena roughly between those that are low-level and those that are high-level, according to whether or not they involve the imagination. I argued that empathy is a process, not an outcome, and I introduced the notion of empathy as it appears in relation to the mental simulation debate in the philosophy of mind. I identified two types of high-level empathy—in-his-shoes imagining and imagining being the other person—and argued that high-level empathy can be spontaneous and deliberate. I also argued that it is a capacity or skill and not a moral virtue. With this overview of LLE and HLE in mind, I now apply the motivation objection to three constitutive views of empathy’s role in morality, beginning in the next chapter with Hume’s sympathy-based moral sentimentalism.
PART I

CONSTITUTIVE VIEWS
2 Hume’s Sympathy

2.1 Introduction

In this chapter I argue that Hume faces the motivation objection. This is my main objection to the constitutive views I will be examining in this part of the thesis. Hume, Smith and Slote hold that empathy’s ethical value lies in the fact that it plays a constitutive role in morality. The motivation objection has three aspects: (1) reliability: because it can be affected and influenced by our motivations, empathy processes are not reliable insofar as they do not always lead to the right results; (2) function: because it can be affected and influenced by our motivations, empathy is not sufficient for various functions, for example causing compassion; (3) circularity: and because it can be affected and influenced by our motivations, using it to define or explain certain aspects of morality in a non-circular way is problematic. In this chapter I argue that Hume faces aspect (2).

Hume has a broad conception of sympathy. It is, on his view, a mechanism for the transmission of sentiments. It plays a vital role in his theory of the passions, particularly in the genesis of the indirect passions. It also plays a fundamental role in his theory of moral judgement, producing the moral sentiments of approval and disapproval. Furthermore, the term sympathy refers to two outputs of the sympathy mechanism, viz. limited sympathy and extensive sympathy. Hume claims that sympathy plays a constitutive role in morality insofar as it is necessary and sufficient for the production of certain sentiments, including compassion and the moral sentiments, and for moral judgement, which, on Hume’s view, consists in feeling the moral sentiments. I will be concerned with aspect (2) of the motivation objection in this chapter: I argue that Hume’s claim that sympathy is sufficient for compassion is false, because our dispositions and characters can affect the outcome of the
sympathizing. I also apply this aspect of the motivation objection to Hume’s theory of moral evaluation, only in this case, his claim is not that sympathy is sufficient for compassion, but rather that it is sufficient for morality. I argue firstly that our dispositions and characters mean that we can fail to make genuine moral judgements, and secondly that Hume’s account of moral judgement entails that no-one would ever act morally.

The chapter is structured as follows. Section 2.2 introduces Hume’s notion of sympathy. It is divided into two parts: the first sets out Hume’s theory of the passions; the second provides exegesis and analysis of the sympathy mechanism. At first blush, it seems like sympathy is akin to emotional contagion or some form of low-level empathy. And just this has been argued for by several commentators, who claim that Hume was an early exponent of some form of low-level simulation.19 However, I argue that Hume’s sympathy concept includes a limited form of HLE that involves the imagination.

Section 2.3 contains exegesis and analysis of sympathy’s role in the production of compassion and benevolence. I apply aspect (2) of the motivation objection to Hume’s claim that sympathy is sufficient for compassion. I argue that imaginative sympathy, as I call Hume’s limited form of HLE, is necessary for compassion, but is not sufficient for it. It cannot guarantee that compassion is produced when we sympathize because what we feel as a result of sympathizing turns out to depend on the motivations—the disposition and character—of the sympathizer.

---

19 See, for example, Collier (2010), Gordon (1995) and Goldman (2006).
Section 2.4 applies aspect (2) of the motivation objection to Hume’s theory of moral judgement. Here, however, the claim is not that sympathy is insufficient for compassion. Rather, it is that sympathy is not sufficient for moral judgement, because the motives of the sympathizer mean that sympathizing cannot guarantee that, if we sympathize, we will make a genuine moral judgement. Moreover, Hume fails to show why we should privilege the moral judgements we do make from the point of view required for moral judgement over our own sentiments, felt as a result of sympathizing from our own point of view. Given this, it is not clear that Hume’s theory possesses the resources to show why anyone would be motivated to be moral. I also defend the view that Hume is consistent in his use of sympathy throughout *A Treatise of Human Nature*. (Hume, 1739/1978)\textsuperscript{20}

In addition, Section 2.4 will serve as a useful introduction to how the notion of sympathy can function in a sentimentalist theory of moral evaluation, different versions of which I consider in Chapters 3 and 4.

### 2.2 Sympathy

#### 2.2.1 The Passions

Hume introduces sympathy in Book II of the *Treatise* to explain how the feelings and opinions of others come to affect and influence our own passions. Sympathy explains, for example, how another person’s approval or disapproval of us can enhance or undermine the passion of pride, and how another person’s suffering can cause pity and compassion.\textsuperscript{21} In this section I give an overview of Hume’s theory of the passions, in addition to the sympathy mechanism. An understanding of the passions will require

\textsuperscript{20} Henceforth abbreviated to ‘*Treatise*’. Citations include page references, and section numbers where these are appropriate.

\textsuperscript{21} For Hume, pity is the same as compassion. The section of the *Treatise* in which he first introduces pity is entitled ‘Of Compassion’ (*Treatise*, p. 368) and it is clear that these two are equivalent from what he says elsewhere (e.g. *Treatise*, pp. 593-594) where he uses the terms ‘pity’ and ‘compassion’ interchangeably. Hume himself usually refers to ‘pity’, but I will be referring to ‘compassion’ in what follows.
explication of some aspects of Hume’s theory of mind, notably impressions and ideas, and the so-called ‘double relation’ between impressions and ideas.

2.2.1.1 Ideas and Impressions

In Book I, Part I, Section I of the *Treatise*, Hume claims that all human perceptions are made up of impressions and ideas. If ideas can best be regarded as thoughts, then impressions are perhaps most usefully described as feelings. (*Treatise*, p. 2) The primary difference between ideas and impressions lies in the ‘degree of force and vivacity’ with which they ‘strike upon the mind, and make their way into our thought and consciousness’. (*Treatise*, p. 1) Both ideas and impressions can be sub-divided into the simple and the complex. *Simple* ideas or impressions cannot be further separated into parts, whereas *complex* ideas or impressions have more than one constituent idea or impression.

Ideas derive from impressions, of which they are weak copies. They differ only in the degree of force, vivacity and liveliness with which they strike upon the mind. (*Treatise*, p. 96) Impressions are divided into original and secondary impressions or, what they are also called, impressions of sensation and impressions of reflection. Original impressions arise ‘in the soul originally, from unknown causes.’ (*Treatise*, p. 7) They ‘arise immediately from nature’ and are the result of natural and physical causes. (*Treatise*, p. 276) They comprise bodily pains and pleasures and impressions of the senses. (*Treatise*, p. 275)

Original impressions give rise to secondary impressions, either directly or ‘by the interposition of [a] corresponding idea.’ (*Treatise*, p. 275) So a fit of the gout causes pain (an original impression) which then produces secondary impressions either directly—I am having a fit of gout right now and it is
painful—or if I contemplate the pain a fit of the gout causes. This would be the interposition of the corresponding idea of the original impression of pain. (*Treatise*, p. 276)

Secondary impressions comprise the passions and emotions. Hume subdivides these in two ways. Firstly, he makes a distinction between *calm* and *violent* secondary impressions. Those described as calm are: ‘the sense of beauty and deformity in action, composition, and external objects.’ (*Treatise*, p. 276) Those described as violent are: ‘the passion of love and hatred, grief and joy, pride and humility.’ (*Ibid.*) This division, he admits, is not an exact one. He observes that sometimes the calm appreciation of a piece of music can send us into violent raptures. (*Ibid.*) Conversely, violent passions ‘may decay into so soft an emotion, as to become, in a manner, imperceptible.’ (*Ibid.*) Hume adopts the distinction because, he argues, it is usually the case that the passions are more violent than the emotions that arise from beauty and conformity, and this is how we usually go about distinguishing between them. And given that it is the violent passions he is interested in explaining in Book II, he adopts the distinction for ease of exposition. The distinction is, however, by no means inflexible. (*Ibid.*)

The second distinction is made between *direct* and *indirect* passions. Some direct and indirect passions can be violent; others are usually considered to be calm. The direct passions are caused by the immediate experience of, or the thought of, good or evil, pleasure or pain. (*Treatise*, p. 276) They include desire, hope, fear, grief, joy and aversion, despair and security. A fit of the gout can cause the direct passions of grief, hope and fear when we either have a fit of the gout, and it is painful, or when we get an idea of the pain gout causes. Certain or probable evil causes grief or sorrow, whereas certain or probable good causes joy. If we think evil is certain, we feel fear.
Similarly, if we think evil is uncertain or unlikely, we can feel hope. (*Treatise*, pp. 439, 574)

The indirect passions, too, arise from the experience or prospect of good, evil, pleasure or pain, but in conjunction with ‘other qualities’. (*Treatise*, p. 276) This will become clearer in the next section. Indirect passions include pride, humility, love and hatred. They arise not only as a result of their connection with entities and events that cause pleasure and pain, but also in virtue of their relationship with ‘that succession of related ideas and impressions, of which we have an intimate memory and consciousness,’ namely our idea of self. (*Treatise*, p. 277)

*Figure 1* contains a basic taxonomy of the passions. The diagram is not meant to suggest any causal connections between the boxes; it is merely intended to illustrate the divisions and distinctions Hume makes. *Figure 2* summarises secondary impressions into calm, violent, direct and indirect.

---

22 Hume is adamant that both the direct and indirect passions are dependent on the presence or prospect of pain or pleasure. Without these, the passions cease to be felt. (*Treatise*, p. 438)

23 There is disagreement among commentators about the consistency of Hume’s account of the self in Books I and II of the *Treatise*, particularly in relation to pride. Some claim that his account is inconsistent between Books I and II, and that this is damaging to his theory of the passions. These issues are important but I will not be discussing them here. For a discussion and defence of Hume against the charge of inconsistency, see e.g. Mercer (1972, p. 30). For a classic discussion of Hume’s theory of pride, see Davidson (1976). Also see Baier (1978). For a modern interpretation, see Martin (2006). For a more detailed discussion of Hume’s conception of the self, see Norton (1982, p. 140) and Mercer (1972, p. 130).
Figure 1: The Passions

<table>
<thead>
<tr>
<th>SECONDARY IMPRESSIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct passions</strong></td>
</tr>
<tr>
<td>Desire, aversion, grief, joy, hope, fear, despair, security, benevolence, resentment, ‘hunger, lust, and a few other bodily appetites’ (T 439)</td>
</tr>
<tr>
<td><strong>Indirect passions</strong></td>
</tr>
<tr>
<td>Pride, humility, love, hatred, pity, malice, ambition, vanity, envy, generosity</td>
</tr>
<tr>
<td><strong>Calm passions</strong></td>
</tr>
<tr>
<td>Sense of beauty and deformity in action, composition, and external objects</td>
</tr>
<tr>
<td><strong>Violent passions</strong></td>
</tr>
<tr>
<td>Love, hatred, grief, joy, pride, humility</td>
</tr>
</tbody>
</table>

Figure 2: Secondary Impressions

Ideas and impressions, according to Hume, differ only in the degree of force, vivacity and liveliness with which they strike upon the mind. (Treatise, p. 96)

This difference can be diminished or strengthened by the famous *principles of association*, which govern the way in which all the perceptions of the mind—the ideas and impressions—interrelate. Although there is no *necessary* connection between them (Treatise, p. 400), certain ideas occur with apparent
regularity alongside certain impressions, and certain impressions often follow on from other impressions and ideas. The principles of association enable Hume’s so-called ‘double relation of impressions and ideas’ to produce the indirect passions, which I explain below, and for these to give rise to other passions and ideas. They can do this, Hume argues, because ‘nature has bestow’d a kind of attraction on certain impressions and ideas, by which one of them, upon its appearance, naturally introduces its correlative.’ (Treatise, p. 289)

The principles of association are the three relations of resemblance, contiguity (in time and space), and causation. Ideas are related by contiguity, causation and resemblance, and impressions are related by resemblance. These three relations, as well as explaining the way in which ideas and impressions tend to interrelate, also facilitate the transmission of force, liveliness and vivacity between ideas and impressions: relations can strengthen and enliven ideas to the extent that they become indistinguishable from their related impressions. (Treatise, p. 319) Moreover, because ideas and impressions differ only in terms of this force and vivacity, if the requisite relations are present, an idea can be converted into an impression. This occurs in the sympathy mechanism, as we will see below. (See Treatise, p. 320)

2.2.1.2 Pride and the Double Relation of Impressions and Ideas

Pride and the other indirect passions are caused by a ‘double relation’ between impressions and ideas. (Treatise, pp. 286, 439) The indirect passions require the experience or thought of pleasure and pain as well as ‘other

---

24 See e.g. Treatise (pp. 10-15) and the Abstract of the Treatise.
25 See Treatise (p. 283). Although later on in the Treatise, Hume introduces another relation that can hold between impressions – directionality. This is crucial for his account of compassion, as we will see below in §2.3. (Treatise, pp. 303-309)
qualities’. (Treatise, p. 279) In Book II, Part I, Section 2 of the Treatise, Hume explains what these qualities are. Each particular indirect passion has an object, a sensation and a cause. The object of pride and humility is always the self: when we feel either pride or humility, it is always with relation to the idea we have of ourselves.\(^{26}\) (Treatise, pp. 277, 286, 574) The idea of self is vividly present and with us at all times. The characteristic sensations of pride and humility are pleasure and pain, respectively. The characteristic sensation is how pride and humility always feel. (Treatise, p. 286) When this sensation is removed, the passion is no longer being experienced.

Pride and humility can have many causes, which can include character traits, physical characteristics, and ‘whatever objects are in the least ally’d or related to us [...] our country, family, children, relations, riches, houses, gardens, horses, dogs, cloaths.’ (Treatise, p. 279) All these causes are subdivided into quality and subject. The quality is the cause’s tendency to produce pleasure or pain; the subject is that in which this quality inheres. The cause must be related to us if it is to result in our feeling pride. (Treatise, p. 288)\(^{27}\)

Our virtues and vices—our good and bad character traits—can also cause pride and humility by the double relation. On Hume’s view, certain characteristics and passions in others and ourselves naturally cause us to feel pain or pleasure ‘by the very view and contemplation.’ (Treatise, p. 296) The

\(^{26}\) The object of love and hatred is usually another person, not the self.  
\(^{27}\) Hume argues that the object of the indirect passions is the thing (idea) that these passions, when they are excited, turn our attention towards. The object of pride and humility is the self; the object of love and hatred is the other person. (Treatise, pp. 277-279) The difference between the direct and indirect passions is that the latter have objects (self or other) that differ from their causes (house, riches, poverty etc). See Cohon (2008a) for a discussion of the distinction between the indirect and direct passions and their causes and objects.
pleasure and pain we feel as a result ‘are not only inseparable from vice and virtue, but constitute their very nature and essence.’ (Treatise, p. 296)

To approve of a character is to feel an original delight upon its appearance. To disapprove of it is to be sensible of an uneasiness. The pain and pleasure, therefore, being the primary causes of vice and virtue, must also be the causes of all their effects, and consequently of pride and humility, which are the unavoidable attendants of that distinction. (Treatise, p. 296)

Hence, if I am aware that I possess some virtuous character trait, I feel pleasure when I contemplate it in myself, and am proud of myself. Pride is summed up as ‘that agreeable impression, which arises in the mind, when the view either of our virtue, beauty, riches or power makes us satisfy’d with ourselves.’ (Treatise, p. 297) Its opposite, humility, is a disagreeable impression that arises upon contemplation of our own vice, deformity, poverty or impotence.

In the case of pride, the double relation works as follows. The idea a person has of his own house, for example, is related to his idea of self. This constitutes the relation of ideas. The characteristic feeling of pride, pleasure, is the impression that is related to the impression of pleasure that the (beautiful) house causes when it is contemplated. The double relation of impressions and ideas is depicted in Figure 3 below. The circle represents John’s mind, and the heart is the passion of pride.
As well as being necessary for pride and humility, the double relation is also necessary for the production of the other indirect passions, including love, hatred and humility. To give rise to humility, John’s house (cause) would have to be unpleasant to look at. It would cause pain upon contemplation of it, and this, together with humility’s characteristic feeling of pain, would constitute the relation of impressions. The relation of ideas would remain the same. The production of love and hatred, on the other hand, would require a change in the relation of ideas because the object of these passions is another person, not ourselves. (*Treatise*, p. 329) Below is a table summarising these four indirect passions, their characteristic sensations, and their objects. I have also included compassion and malice, two other indirect passions.
2.2.2 The Sympathy Mechanism

With this overview of ideas, impressions and the passions in mind, we are now in a position to see what Hume means by sympathy. Sympathy is first introduced in Book II, Part I, Section 11 of the Treatise as a secondary cause of pride and humility, when these passions are affected by the positive or negative sentiments and opinions of other people. (Treatise, pp. 316-324, ‘Of the love of fame’) For example, if John’s house causes him to feel proud, then the admiration of another person can further strengthen and enliven his pride via sympathy. Hume considered sympathy to be the fundamental principle of human morality, and it appears throughout Books II and III of the Treatise, as well as in the second Enquiry (Hume, 1777/1975).  

---

28 See e.g. Treatise (p. 618). It is a vexed question in Hume scholarship whether or not Hume abandons the Treatise notion of sympathy in the second Enquiry. Remy Debes (2007b) calls this ‘the puzzle of Hume’s second enquiry’. This is the claim that Hume, after having taken pains to argue for and outline the sympathy mechanism and its role in the production of the moral sentiments in the Treatise, seems to abandon talk of it altogether in the second Enquiry. Indeed, he appears to
No quality of human nature is more remarkable, both in itself and in its consequences, than that propensity we have to sympathize with others, and to receive by communication their inclinations and sentiments, however different from, or even contrary to, our own. (Treatise, p. 316)

Sympathy does a lot of work in the Treatise, not only in Book II, but also in Book III. It is responsible for explaining how we are affected by what others feel and how we come to be influenced by those feelings. (Treatise, p. 317) Sympathy is not a passion. Rather, it is a causal mechanism that explains how the sentiments and opinions of others come to influence and affect us. It is a process consisting of the conversion of the idea we have of the passion another person is experiencing into ‘the very passion’ in ourselves. (Treatise, p. 317) It operates in the same manner as the understanding, that is, by the principles of association. (Treatise, pp. 319-320) We always have a lively idea of ourselves in our minds. Whatever objects are related to us are ‘conceived with a like vivacity of conception.’ (Treatise, p. 317) Other people are related to us because they resemble us. The more they resemble us—in appearance, character, nationality, ethnicity and so on—the stronger this relation will be. Other people are also related to us by contiguity: the closer another person is to us in time and space, the stronger this relation will be.33

replace it with the ‘principle of humanity’. I will not enter this debate here, as I am primarily concerned with sympathy in the Treatise in this thesis. However, see Debes (2007b) for a convincing argument in favour of Hume’s consistency in the Treatise and the second Enquiry, and his (2007a) for an equally convincing argument for the claim that Hume does not abandon his associationist account of sympathy, outlined at such length in the Treatise, in the second Enquiry.35 Cf. Mercer (1972, p. 26), Penelhum (1975, p. 147) and Stroud (1977, p. 196).

30 Hume sometimes refers to ‘the communicated passion of sympathy.’ (Treatise, pp. 370, 576) I take him to mean the passion we feel as a result of sympathizing, i.e. the passion of the other person that we now feel as a result of converting an idea of it into an impression in ourselves.


32 We can take the imagination to be synonymous with the understanding: ‘The imagination or understanding, call it which you please.’ (Treatise, p. 440)

33 It is not made clear in the Treatise how Hume thinks we can or do come to know about the existence of other selves, which we must know about if we are to know that they have sentiments, and are psychologically similar to us. See Pitson (1996) for a discussion of the self and its relation to sympathy.
Hume denies that we can have access to the sentiments of others in any other way than by inference from effect to cause or cause to effect. (*Treatise*, pp. 320, 576) This is because ‘no passion of another discovers itself immediately to the mind. We are only sensible of its causes and effects. From these we infer the passion: And consequently these give rise to our sympathy.’ (*Treatise*, p. 576) It is not the case that people’s *occurrent* sentiments must determine the initial idea we obtain of their passion; our initial idea depends rather on an inference from cause to effect or effect to cause. We can either infer from the usual effect of a passion (a smile, for example) to its usual cause (pleasure, for example), or from its usual cause (poverty, for example) to its usual effect (misery, for example).34

When we see the effects of a passion manifested in ‘external signs in the countenance and conversation’ (*Treatise*, p. 317) of a person, we get an idea of the cause, which is the passion they are feeling. We can also get our idea of a passion by inferring from something we believe usually causes it to its usual effect, the passion. (*Treatise*, p. 576) If, for example, we see instruments of surgery in front of us, we infer from these causes to their usual effects, and obtain related ideas of pain and suffering. (*Treatise*, p. 576)

> When I see the *effects* of a passion in the voice and gesture of any person, my mind immediately passes from these effects to their causes, and forms such a lively idea of the passion, as is presently converted into the passion itself. In like manner, when I perceive the *causes* of any emotion, my mind is convey’d to the effects, and is actuated with a like emotion. (*Treatise*, p. 576, Hume’s italics)

Once we have an idea of the passion, the relations of resemblance and contiguity strengthen and enliven it to such an extent that it is converted into its corresponding impression: the only thing that distinguishes between

---

34 These causal inferences are based on custom and habit in the form of general rules. This will be important later on.
ideas and impressions is their different degrees of force and vivacity. The relations, as we have seen, are capable of transmitting this force and vivacity from an idea to an impression until they become indistinguishable from each other. (Treatise, p. 319) When the idea becomes the impression, it has become ‘the passion itself’ in the sympathizer. Below are the steps A undergoes when sympathizing with B.

1. A always has a lively idea of himself
2. B resembles and is contiguous with A
3. A sees B smile
4. A infers from B’s smile (effect) to its usual cause (pleasure)
5. A obtains an idea of B’s pleasure
6. Contiguity and resemblance enliven A’s idea of B’s pleasure by relating it to his idea of self
7. A’s idea of B’s pleasure becomes A’s impression of pleasure

Returning briefly to sympathy’s role in the production of pride, we can see that it operates as follows. If John feels proud of his house, this pride can be enhanced by Jane’s approval of it. This is because John becomes aware of—and influenced by—Jane’s positive opinion, which happens in the following way.

Now nothing is more natural than for us to embrace the opinions of others in this particular; both from sympathy, which renders all their sentiments intimately present to us; and from reasoning, which makes us regard their judgement, as a kind of argument for what they affirm. These two principles of authority and sympathy influence almost all our opinions. (Treatise, p. 321, Hume’s italics)
So John gets an idea of Jane’s pleasure—his pleasurable sentiment of approval—which sympathy then converts into an impression of pleasure in him. This reinforces John’s own feeling of pride.\(^{35}\)

2.2.2.1 Sympathy: Low- or High-level?

At first blush, Hume’s sympathy looks like some sort of low-level empathy (LLE) process like emotional contagion or mirroring. This has been argued for by, for example, Slote (2010a), Goldman (2006), Gordon (1995) and Mark Collier (2010). Collier, for example, claims that Hume’s sympathy is a contagion-like view based on mirror neurons. (Collier, 2010, p. 255)\(^{36}\)

A contagion view is often attributed to Hume based on interpretation of three of his metaphorical descriptions of sympathy, two from the Treatise and one from the second Enquiry.\(^{37}\) I will quote these below, plus a further quotation from the Enquiry that also appears to lend support to the idea that sympathy is a form of LLE.

In general we may remark, that the minds of men are mirrors to one another, not only because they reflect each other’s emotions, but also because those rays of passions, sentiments and opinions may be often reverberated, and may decay away by insensible degrees. (Treatise, p. 365)

We may begin by considering a-new the nature and force of sympathy. The minds of all men are similar in their feelings and operations, nor can any one be actuated by

\(^{35}\) Add to this the stipulations that we are always partial in our own favour (Treatise, p. 321), and that we care more about the opinions of those we ourselves respect (usually because they bear a close relation to us), and the result is that if someone whose opinion John respects admires a house that he himself is proud of, and is in actual fact beautiful, his impression of Jane’s pleasure, obtained via sympathy, plus his own pleasure caused by the beauty of his house, gives him an added feeling of pleasure. If, however, A is proud of his house and someone whose opinion he respects disagrees, he will feel pain: ‘the uneasiness of being contemn’d depends on sympathy.’ (Treatise, p. 322) This painful feeling will input into the double relation and give rise to humility. If A does not respect the opinion of a person who disapproves of his house, or if he is very proud by nature, the opinion of the other person will not diminish his pride, and he will not be ashamed.

\(^{36}\) Collier also attributes what he calls a ‘cognitive’ sympathy to Hume, which is similar to a kind of high-level empathy, about which I will be saying more below.

\(^{37}\) See e.g. Collier (2010, pp. 256-257) and Debes (2007, p. 45).
any affection, of which all others are not, in some degree, susceptible. As in strings equally wound up, the motion of one communicates itself to the rest; so all the affections readily pass from one person to another, and beget correspondent movements in every human creature. (*Treatise*, p. 576, Hume’s italics)

The human countenance, says Horace, borrows smiles or tears from the human countenance. (Hume, 1777/1975, p. 220)

A man who enters the theatre, is immediately struck with the view of so great a multitude, participating of one common amusement; and experiences, from their very aspect, a superior sensibility or disposition of being affected with every sentiment, which he shares with his fellow-creatures. (Hume, 1777/1975, p. 221)

What Hume has to say about animals’ ability to sympathize also points towards sympathy being some sort of LLE process. He claims that animals have a sense of each other’s pains and pleasures (*Treatise*, p. 398), and that sympathy is observable ‘through the whole animal creation’. (*Treatise*, p. 363)

Sympathy, or the communication of passions, takes place among animals, no less than among men. Fear, anger, courage and other affections are frequently communicated from one animal to another, without their knowledge of that cause, which produc’d the original passion. Grief likewise is receiv’d by sympathy; and produces almost all the same consequences, and excites the same emotions as in our species. (*Treatise*, p. 397)

Given all this evidence, it is easy to jump to the conclusion that Hume’s sympathy is a form of LLE. I do not want to deny that animals can undergo a communication of sentiments in a low-level way, or that humans might be able to sympathize in a low-level way with direct passions like grief. However, I suggest that there is another type of sympathy at work when it comes to the moral sentiments and certain passions in humans, a type of sympathy that involves the imagination, the capacity for which animals do not possess. I argue that Hume invokes this type of sympathy, and that it is
necessary for compassion. And we will see that, if sympathy were LLE, it would not be possible for it to produce the outcome, in this case what Hume calls ‘extensive sympathy’, which is necessary for pity. Before doing this, however, I will argue that animal sympathy cannot be the same as human sympathy, at least in some cases. This implies that there are at least two forms of sympathy at work in Hume.

In the parts of the Treatise where he discusses the indirect passions, Hume has sections at the end of each one devoted to the passions of animals. Animals can feel pride, humility, love and hatred, and the same things that cause these passions in humans can cause them in animals, too. But animals do not feel these passions as a result of the double relation. Rather, they do so because of the causes themselves and the pleasure or pain these cause directly. So if a dog’s master feeds it, it will love him because he caused it pleasure by doing so; if he beats it, it will hate him because he caused it pain. (Treatise, p. 397) Thus the cause and object of the passion are the same: the master.

In humans, by contrast, love and hatred have different causes and objects. The love and hatred we bear towards others is determined not only by the power of the cause to produce pleasure and pain in us, but also by various other relations, as outlined earlier. Although animals can ‘trace relations’ between objects, they cannot do this ‘except in very obvious circumstances.’ (Treatise, p. 397) Animals cannot trace relations as we can because they have inferior reason and understanding. (Treatise, pp. 326, 610) They can love their master, and members of their own species, for example, because of a basic

---

38 See Treatise (pp. 324-328, 397-398).
39 Hume also says that animals cannot feel the moral sentiments because they have ‘little sense of virtue or vice.’ (Treatise, pp. 326, 468)
relationship of acquaintance. But they cannot, as we can, love (esteem) someone when the object and cause are different, for example when we love a person (object) on account of something that he possesses (cause). We love and esteem the rich, on Hume’s view, because we sympathize with—feel by communication—the pleasure that their riches cause them to feel. (Treatise, p. 321) To sympathize in this case requires knowledge of what caused the object to feel pleasure, viz. the riches.

Another reason to suppose that sympathy can operate differently in humans to produce certain passions can be demonstrated by the fact that some passions, specifically those that depend on the imagination, are not readily felt by animals. In humans, compassion depends entirely on the imagination and on sympathy, as we will see below in §2.3. (Treatise, p. 371) But it is clear from the following two quotations that animals do not have the imaginative capacities that we do, which points towards their being unable to sympathize as we can, given that sympathy, at least in the case of compassion, requires the imagination.

As animals are but little susceptible either of the pleasures or pains of the imagination, they can judge of objects only by the sensible good or evil, which they produce, and from that must regulate their affections towards them. (Treatise, p. 397, Hume’s italics)

Envy and malice are passions very remarkable in animals. They are perhaps more common than pity; as requiring less effort of thought and imagination. (Treatise, p. 398)

The brevity of the above discussion notwithstanding, I hope to have demonstrated that, at the very least, animals and humans can sympathize differently, so it is not straightforward to jump from the evidence that sympathy exists in animals, and looks as though it might be low-level, to the
conclusion that human sympathy is only low-level as well. Animal sympathy *cannot* be the same as human sympathy, at least in some cases. One such case is when it gives rise to compassion. For, as I argue in the following section, compassion depends on what I will call *imaginative sympathy*.

Imaginative sympathy is different from the two forms of HLE I described in Chapter 1—imagining being the other person and in-his-shoes imagining—because it does not involve adopting the perspective of another person or putting oneself in their shoes. But it does involve the imagination, which makes it high-level, according to my distinction. It must therefore be distinguished from the low-level, animal form of sympathy. As I will argue in the next section, Hume relies on imaginative sympathy to deal with cases in which the other person is not feeling a sentiment, or is feeling a different sentiment to the one he ought to be feeling, given his circumstances. In these cases, imaginative sympathy enables him to imagine, given the other person’s present condition (cause), what effect (passion) it usually gives rise to. Or, alternatively, to imagine, given some effect of a passion, what the usual cause of that effect is. Because the imagination is what enables the sympathizer to make the initial causal inference necessary for sympathy, and this causal inference is necessary for paradigmatic cases of compassion, imaginative sympathy is necessary for compassion. A contagion-like, low-level sympathy would not be able to facilitate the making of such causal inferences. And low-level mechanisms of this kind depend on the person who is feeling the passion (the target) being nearby, so the sympathizer can ‘pick up’ his passion, and depend on the target actually feeling the passion that the sympathizer picks up. Hume has a variety of examples of cases in which the target is *not* feeling a sentiment, or is feeling a *different* sentiment,
which means he cannot be relying on a form of sympathy that could not deal with these more complex cases.

2.3 Compassion and Benevolence

Compassion requires the imagination, and Hume himself says just this. It also depends on sympathy. In this section I apply the motivation objection to Hume’s claim that sympathy is sufficient for compassion. Before doing so, however, I explain how sympathy gives rise to compassion and the related passion of benevolence. I then argue that imaginative sympathy is necessary for compassion by citing examples Hume provides in the Treatise. Finally, I apply aspect (2) of the motivation objection: I argue that Hume fails to show that sympathy is sufficient for one of its functions, in this case for compassion. What will turn out to be most interesting about this, however, is that Hume appears to set himself up for the objection in Book I of the Treatise.

2.3.1 The Production of Compassion and Benevolence

Compassion and benevolence, as described in the Treatise in Book II, are different passions. Compassion is an indirect passion, consisting of a desire for the happiness and aversion to the misery of strangers, and a characteristic painful feeling. It is a ‘concern for [...] the misery of others, without any friendship [...] to occasion this concern.’ (Treatise, p. 369) The object of the passion is the other person, and its cause is whatever caused that person’s misery. Benevolence is a calm, direct passion, comprising a desire for the happiness, and aversion to the misery, of those we already bear some connection to—for example, our friends or family—and has a ‘warm’ or tender sensation. (Treatise, p. 367)
We always feel benevolence for those we love, and anger for those we hate. This is because love and hatred are ‘always followed by, or rather conjoin’d with benevolence and anger,’ which means that we ‘automatically’ feel them when we love or hate a person. (*Treatise*, p. 367) But the connection between love and benevolence is not necessary, Hume claims; the fact that benevolence always appears with love should not be taken to imply that it must only arise with it. It is ‘an original pleasure arising from the pleasure of the person below’d, and a pain proceeding from his pain: From which correspondence of impressions there arises a subsequent desire of his pleasure, and aversion to his pain’. (*Treatise*, p. 387) But benevolence can also ‘frequently arise from a natural impulse or instinct, which is perfectly unaccountable.’ (*Treatise*, p. 439) This kind, also described as ‘happiness to our friends’ when it appears with love (*Treatise*, p. 417), is an instinct originally implanted in our nature. (*Treatise*, p. 439) Hume also has a place for benevolence that is not limited to our family and friends: we can feel benevolence for people in general, as will be illustrated by Example 5 in §2.3.2 below. However, Hume argues that there is ‘no such passion in human minds, as the love of mankind, merely as such, independent of personal qualities, of services, or of relation to oneself.’ (*Treatise*, p. 481) Benevolence for our friends arises because it accompanies love; benevolence for strangers arises as a result of the sympathy mechanism. (*Treatise*, p. 481)

Benevolence, ‘properly speaking, produce[s] good and evil, and proceed[s] not from them.’ (*Treatise*, p. 439) And benevolence is among a number of ‘calm desires and tendencies, which, tho’ they be real passions, produce little emotion in the mind, and are more known by their effects than by the immediate feeling or sensation.’ (*Treatise*, p. 417) So part of having

---

40 Anger is the opposite passion to benevolence, viz. a desire for the misery of those we hate, and aversion to their happiness. (*Treatise*, p. 367)
benevolence for another person involves desiring their good and being averse to their misery. It is plausible to suggest that the effect of benevolence is acting to help the person we have benevolence for—benevolent motivation, in other words—or to prevent harm to him.\textsuperscript{41}

Compassion ‘imitates the effects’ of benevolence, insofar as it can prompt benevolent action, but is only caused by the misery of others, and is a painful impression. (\textit{Treatise}, p. 372) It is always \textit{associated} with benevolence and love, but is different from them phenomenologically because it is a painful feeling rather than a warm and tender one. And benevolence is a direct passion, whereas compassion is indirect.\textsuperscript{42}

Hume claims that ‘there is always a mixture of love or tenderness with compassion, and of hatred and anger with malice.’ (\textit{Treatise}, p. 381) But this seems to be contradictory to his general theory of the passions, according to which impressions are related to each other by resemblance, which explains why they appear together and give rise to each other. (See \textit{Treatise}, p. 381) Hume argues at length in the \textit{Treatise} that we feel contempt for the poor, rather than love and tenderness, because we sympathize with the misery that their poverty causes them to feel. Misery and contempt are both painful passions, so the disagreeable feeling the poverty causes gives rise to our

\textsuperscript{41}Interestingly, Hume speaks of something like benevolence existing in animals: ‘The howlings and lamentations of a dog produce a sensible concern in his fellows. And ‘tis remarkable, that tho’ almost all animals use in play the same member, and nearly the same action as in fighting; a lion, a tyger, a cat their paws; an ox his horns; a dog his teeth; a horse his heels: Yet they most carefully avoid harming their companion, even tho’ they have nothing to fear from his resentment; which is an evident proof of the sense brutes have of each other’s pain and pleasure.’ (\textit{Treatise}, p. 398)

\textsuperscript{42}In §II.i.9 of the \textit{Treatise} (pp. 438-448), entitled ‘Of the direct passions’, Hume says: ‘desire arises from good consider’d simply, and aversion is deriv’d from evil. The will exerts itself, when either the good or the absence of the evil may be attain’d by any action of the mind or body.’ (\textit{Treatise}, p. 439) Hume also refers to good and evil as synonymous with pleasure and pain in this section. In the case of extensive sympathy, about which I will be saying more below (§2.3.1.1), an idea of another person’s pain produces two direct passions: a desire for a person’s good and an aversion to his misery. I consider this double impression to be what constitutes extensive sympathy. Hume refers to volition and desire as impressions on the same page. (\textit{Treatise}, p. 439)
hatred and contempt because these impressions are related by resemblance. (Treatise, p. 362) According to Hume’s system, therefore, it ought not to be the case that compassion—because it is a painful feeling—is associated with warm feelings like love. Rather, compassion should, by Hume’s own lights, produce or mix with hatred because it resembles it. (Treatise, p. 381) Hume resolves this problem with what we can call his ‘Chain of the Passions’ (Treatise, p. 382), depicted in Figure 5 below.

![Figure 5: Hume’s ‘Chain of the Passions’](image)

**KEY:**

- - - = Related because both are desires for someone’s happiness/aversions to their misery (parallel direction)
- - - - - = Related by an original and natural connection

---

43 Similarly, we feel love and esteem for the rich because of the pleasure their riches cause in them. (Treatise, pp. 357-365)
44 Esteem and contempt are ‘at bottom the same passions’ as love and hatred, only ‘diversify’d by some causes.’ (Treatise, p. 337)
45 Malice is compassion’s opposite, and Hume claims that malice—a pleasure taken in another person’s misery—can cause anger and hatred. (Treatise, pp. 438-448)
This diagram shows how love, benevolence and compassion are related to each other, and can therefore be associated with each other. As is clear from the Key, Hume invokes a new kind of relation between impressions to explain how a painful feeling of compassion can be associated with warm, tender, pleasant feelings like love and benevolence. (Treatise, p. 384) This new relation is the resemblance of parallel direction. The impressions have a parallel direction because they both include a tendency to desire the happiness, and be averse to the misery, of others. Although it has a different characteristic sensation, compassion is similar to benevolence because it incorporates, like benevolence, an aversion to another person’s misery and a desire for his happiness. Sympathy with someone else’s pain can therefore produce benevolence—a desire for happiness and aversion to misery that naturally accompanies love—because the directions of both these passions are the same.

In the case of pride, it is ‘the present sensation or momentary pain or pleasure, which determines the character of [the] passion.’ (Treatise, p. 384) In the case of compassion, however, it is ‘the general bent or tendency of it from beginning to end.’ (Treatise, p. 384) Some passions, like pride and humility, are ‘pure sensations, without any direction or tendency to action’ so they cannot be related by direction. (Treatise, p. 382) Benevolence and compassion are both examples of passions that do have such a direction or tendency, which means they can be related in this way. Compassion can be associated with benevolence because they both go in the same direction. This means that even though they feel different, they can still resemble each other and thus be associated with one another and appear together.

However, invoking the resemblance of parallel direction does not resolve another potential contradiction in Hume’s system: if the passions can
resemble each other in this new way, would this not mean that another’s poverty always causes benevolence and love, and never contempt? (Treatise, p. 384)46 It is in answer to this conundrum that Hume invokes the distinction between limited and extensive sympathy.47 This is a crucial distinction, not least because extensive sympathy is what goes on, in Book III if the Treatise, to play a vital role in moral evaluation, as we will see in §2.4.

2.3.1.1 Extensive and Limited Sympathy

It is extensive sympathy that causes another’s poverty to give rise to compassion, but limited sympathy which results in their poverty causing contempt. The kind of sympathy that has appeared thus far in the Treatise, for example as a secondary cause of pride and humility, is limited sympathy. The same ‘general principle’ of sympathy applies in the case of compassion. We have a lively idea of everything that is related to us: other people are contiguous to us and resemble us, and are therefore related to us (by resemblance and contiguity), which means that ‘their persons, therefore, their interests, their passions, their pains and pleasures must strike upon us in a lively manner, and produce an emotion similar to the original one; since a lively idea is easily converted into an impression.’ (Treatise, p. 369) However, as the following quotation shows, in the case of compassion, another’s misery causes extensive sympathy, which gives rise to compassion

46 The full quotation is as follows: ‘I have endeavour’d to prove, that power and riches, or poverty and meanness; which give rise to love or hatred, without producing any original pleasure or uneasiness; operate upon us by means of a secondary sensation deriv’d from a sympathy with that pain or satisfaction, which they produce in the person, who possesses them. From a sympathy with his pleasure there arises love; from that with his uneasiness, hatred. But ‘tis a maxim, which I have just now establish’d and which is absolutely necessary to the explication of the phenomena of pity and malice, “That ‘tis not the present sensation or momentary pain or pleasure, which determines the character of any passion, but the general bent or tendency of it from beginning to end.” For this reason, pity or a sympathy with pain produces love, and that because it interests us in the fortunes of others, good or bad, and gives us a secondary sensation correspondent to the primary; in which it has the same influence with love and benevolence. Since then this rule holds good in one case, why does it not prevail throughout, and why does sympathy in uneasiness ever produce any passion beside good-will and kindness?’ (Treatise, p. 384)

47 Hume also calls extensive sympathy ‘compleat’ sympathy and ‘double’ sympathy. (Treatise, pp. 388-389)
and benevolence, by the new kind of relation between impressions, the resemblance of parallel direction. But in the case of contempt, another’s misery causes only limited sympathy, which gives rise to hatred and contempt by the double relation of impressions and ideas.

I have mention’d two different causes, from which a transition of passion may arise, viz. a double relation of ideas and impressions, and what is similar to it, a conformity in the tendency of any two desires, which arise from different principles. Now I assert, that when sympathy with uneasiness is weak, it produces hatred or contempt by the former cause; when strong, it produces love or tenderness by the latter. This is the solution of the foregoing difficulty. (Treatise, p. 385)

Extensive sympathy is a double impression because it is a desire for another’s happiness and aversion to his misery. Extensive sympathy arises when the sympathizer obtains a strong idea of another person’s misery that comes to ‘extend’ over related ideas, giving rise to a concern for the other person’s past, present and future wellbeing. These ideas are related to self, and converted into a desire for the other’s happiness and aversion to his misery. As we are averse to our own misery, and desire our own happiness, by relating the idea of another’s misery to our idea of self, we become concerned for the other person just as we would be for ourselves.\(^{48}\) We come to desire his happiness and are averse to his misery, as we are to our own. This is the same ‘double tendency of the passions’ (Treatise, p. 387) that we find in compassion and benevolence. It can therefore be associated with compassion, benevolence and love (because this always accompanies benevolence).\(^{49}\)

---

\(^{48}\) As Pal Árdal (1966, p. 46) puts it: ‘To to sympathize with x’s desire for brandy is to desire brandy.’ In extensively sympathizing with another person, x desires his own happiness and I desire it too.

\(^{49}\) A strange and surely unwelcome implication of Hume’s insistence on relations obtaining between the passions is that it looks as though poverty would make us want to *harm* the poor, as well as feeling contempt for them. For anger always accompanies hatred because the two are connected ‘naturally’, and anger is a desire for the misery of the person we hate, and aversion to his happiness. So it looks as though we would actively seek the misery of the poor.
Limited sympathy, in the case of contempt, is a single, painful impression. Limited sympathy is caused by an idea of another’s passion—in this case, his misery—that is weak, and does not therefore extend over related ideas. If the initial idea is not lively enough, it cannot diffuse over related ideas, and ‘the future prospect, which is necessary to interest me perfectly in the fortune of another’ is destroyed. (Treatise, p. 386) In this case there is no conversion of the idea into a double impression—only into a single one. The sympathizer can, therefore, only ‘feel the present impression, but carry [his] sympathy no farther, and never transfuse the force of the first conception into [his] ideas of the related objects.’ (Treatise, p. 386) He feels the painful impression, but not the desire/aversion necessary to relate it to compassion and benevolence. If I have sympathized with only a weak idea, ‘I receive it by communication, and am affected with all the passions related to it: But as I am not so much interested as to concern myself in his good fortune, as well as his bad, I never feel the extensive sympathy, or the passions related to it.’ (Treatise, p. 386) The double relation produces hatred or contempt for poverty when a weak idea of another’s misery relates to the idea of self, and the characteristic sensation of hatred (pain or displeasure) relates to the sympathetically communicated (weak, painful) impression of another’s misery. Figure 6 below depicts how limited and extensive sympathy arise, and how they then go on to give rise to the passions associated with them.
Figure 6: Extensive and Limited Sympathy

KEY:

- Related by parallel direction (both are desires for someone’s happiness/aversions to their misery)
- Related by original and natural connection
- = Caused by double relation of impressions and ideas
- = Related by resemblance

CAUSE Severe poverty

EFFECT Severe misery

Strong idea

Extensive sympathy (double impression)

Compassion

Benevolence

Love

CAUSE Slight poverty

EFFECT Slight misery

Weak idea

Limited sympathy (single impression)

Hatred / contempt

Anger

Sympathizing occurs

CAUSE Slight poverty

EFFECT Slight misery

Strong idea

Extensive sympathy (double impression)

Compassion

Benevolence

Love

EFFECT Severe misery

Severe misfortune

EFFECT Slight misfortune
Hume summarizes how sympathizing with a weak, painful impression causes only limited sympathy, but sympathizing with a strong, painful impression causes extensive sympathy, in the following quotation.

When we sympathize only with one impression, and that a painful one, this sympathy is related to anger and to hatred, upon account of the uneasiness it conveys to us. But as the extensive or limited sympathy depends upon the force of the first sympathy; it follows, that the passion of love or hatred depends upon the same principle. A strong impression, when communicated, gives a double tendency of the passions; which is related to benevolence and love by a similarity of direction; however painful the first impression might have been. A weak impression, that is painful, is related to anger and hatred by the resemblance of sensations. Benevolence, therefore, arises from a great degree of misery, or any degree strongly sympathiz’d with; which is the principle I intended to prove and explain. (Treatise, p. 387)

On my interpretation of how limited and extensive sympathy arise, as is evident from the diagram, a weak or strong idea is converted by the sympathy mechanism straight into the two outputs: either limited or extensive sympathy. Rico Vitz (2004), however, takes a different view. Vitz has interpreted this passage as implying an extra stage in the sympathy process, a ‘first sympathy.’ (Vitz, 2004, p. 266) Hume states in the quotation that ‘the extensive or limited sympathy depends upon the force of the first sympathy.’ (Treatise, p. 387) Vitz claims that our initial idea becomes a ‘first sympathy’, which is a weak or strong impression, which then gives rise to limited or extensive sympathy. It is the force of this first impression that determines whether a single or a double impression is then produced.

However, I see no reason to believe that there is an extra, felt impression that, if weak, becomes limited sympathy or, if strong, becomes extensive sympathy. Hume is clear that sympathy is ‘nothing but the conversion of an
idea into an impression by the force of the imagination.’ (Treatise, p. 427) What Vitz proposes would add another stage to the sympathy process, which would mean there would be an initial idea, then a first sympathy (impression), then either limited or extensive sympathy. But why would an extra impression, a ‘first sympathy’, become necessary in the case of limited and extensive sympathy, when it is not required in the case of the sympathetic conversion of the other passions? And why cannot an idea convert into either a single impression or a double impression, depending on the strength or force of the initial idea, and what this initial idea is related with, as is the case in limited and extensive sympathy? In my view, this is what Hume appears to be claiming. He argues that, if the misery is very great, then the idea will be strong. Conversely, when the misery is weak, the idea will be weak. Perhaps the term ‘sympathy’ in ‘first sympathy’ should be taken adverbially: ‘first sympathy’ means the initial conversion of the strong or weak idea into the limited or extensive sympathy. In other words, perhaps ‘first sympathy’ is better read as ‘first sympathizing’. Vitz himself notes (2004, p. 264) that Hume uses the term ‘sympathy’ to refer both to the mechanism and to the communicated passion of sympathy. Although nothing in what follows hinges on this interpretation, it seems preferable to read Hume in the way I suggest, given what else he says about the genesis of limited and extensive sympathy. When we get an initial, strong idea of another’s misery, this is vivified by the imagination by being related to other ideas pertaining to the condition of the other person, which serves to strengthen and enliven this initial idea, to the extent that it becomes a double impression. (Treatise, p. 386) In contrast, when it comes to limited sympathy, the weakness of the initial idea does not ‘transfuse the force of the first conception into my ideas of the related objects,’ so I merely feel the limited sympathy. (Treatise, p. 386) This points to it being more likely that a weak idea converts straight into a weak impression, limited sympathy, and a
strong idea becomes a strong impression, extensive sympathy, and does not imply the existence of an intermediary ‘first sympathy’ impression.

2.3.1.2 Sympathy and Concern

Several commentators have criticised Hume for failing to provide an adequate account of how sympathy can cause the kind of concern for the well-being of others that compassion and benevolence imply. Commentators like Douglas Chismar (1988) and Pal Árdal (1966) have argued that Hume fails to show how simply having another person’s pain communicated to us can make us concerned for their well-being. As Árdal (1966, p. 51) puts it, Hume seems ‘not to be adequately alive to the need to explain why a communication of sorrow or unhappiness from another should occasion in us a concern about the other person.’ In a similar vein, Chismar (1988, p. 244) argues that ‘Hume’s description of what occurs in sympathizing does not contain within it sufficient explanation for the occurrence of a benevolent attitude.’

But Hume does appear, on the face of it, to have an account of how sympathy can cause concern. He claims that the way sympathy causes compassion and benevolence is via the conversion of our strong initial idea of another person’s (great) misery into the double impression of extensive sympathy, which is related to compassion and benevolence by parallel direction. Because extensive sympathy is an aversion to the pain of the other and a desire for his happiness, it relates to benevolence and compassion, both of which include a desire for another’s happiness and aversion to his misery. We can become concerned for the other person because our idea of their passion relates to our idea of self. Hume argues that, if a person’s misery ‘has any strong influence upon me, the vivacity of the conception is not conﬁn’d merely to its immediate object, but diffuses its influence over all
the related ideas, and gives me a lively notion of all the circumstances of that person, whether past, present, or future.’ (Treatise, p. 386) When we sympathize, the idea of the other’s passion is related to our idea of self. This enables the idea of the other and his misery to be diffused over related ideas, generating concern for his well-being. We become concerned for his well-being as we are for our own. So it does appear that Hume can go some way towards explaining how sympathy can cause concern for another’s well-being of the kind that we might associate with benevolence and compassion.

However, for Hume’s response to be plausible we would have to accept his associationism, because it is this that would enable the principles of association to relate the idea of self to the idea of the other’s misery, and extensive sympathy to relate to compassion and benevolence. And there is another potential problem, given that we must also accept Hume’s claim that ‘a certain degree of poverty produces contempt; but a degree beyond causes pity and good-will.’ (Treatise, p. 387) For this is how Hume tackles the possible objection to his view that extensive and limited sympathy is supposed to resolve, namely how another’s misery can cause both compassion and contempt. So Hume needs to show that a certain level of poverty always causes extensive sympathy. In §2.3.3 I argue that he cannot do this: he fails to show that sympathizing with another’s misery guarantees that we feel the extensive sympathy that is necessary for compassion. Before doing so, however, I argue that imaginative sympathy is necessary for compassion.

2.3.2 Imaginative Sympathy and Compassion

I have outlined how the sympathy mechanism operates to produce compassion and benevolence. Here are five examples highlighting the necessity of imaginative sympathy for compassion. They show that
imaginative sympathy is necessary, in each case, to enable the sympathizer to feel compassion based on what they believe the other person does, or would, or should, feel, rather than on what they do feel. This requires an inference from effect to cause or vice versa, which would not be possible if the sympathy in operation here were a contagion-like LLE process.

2.3.2.1 Example 1

This example demonstrates how it is possible for us to feel compassion for another person even though they are not themselves feeling sad or miserable. That is, it demonstrates how compassion can arise ‘by a transition from affections, which have no existence.’ (Treatise, p. 370)

> When a person of merit falls into what is vulgarly esteem’d a great misfortune, we form a notion of his condition; and carrying our fancy from the cause to the usual effect, first conceive a lively idea of his sorrow, and then feel an impression of it, entirely overlooking that greatness of mind, which elevates him above such emotions, or only considering it so far as to increase our admiration, love and tenderness for him. We find from experience, that such a degree of passion is usually connected with such a misfortune; and tho’ there be an exception in the present case, yet the imagination is affected by the general rule, and makes us conceive a lively idea of the passion, or rather feel the passion itself, in the same manner, as if the person were really actuated by it. (Treatise, p. 370, Hume’s italics)

The person of merit is experiencing great misfortune (cause) and this usually causes sorrow (effect). But the man is not himself miserable, for his ‘greatness of mind’ has elevated him above self-pity. However, we still feel sorry for the man even though he is not sorry for himself. We ‘entirely overlook’ his greatness of mind, and form an idea of his misery, the usual effect of a loss of fortune, in the absence of any real effect having been caused. We can see that the imagination is necessary for compassion here because the compassion is arising in spite of the fact that the man is not
feeling miserable. If we sympathized using a contagion-like LLE process, the sympathy would have to be with what the person was actually feeling.

2.3.2.2 Example 2

In this example, the sympathizer is able to sympathize with what he imagines the target would be feeling if he were aware of his own behaviour.

From the same principles we blush for the conduct of those, who behave themselves foolishly before us; and that tho‘ they shew [sic] no sense of shame, nor seem in the least conscious of their folly. All this proceeds from sympathy. (Treatise, p. 371)

The misery or pain the foolish man would feel, if he were truly cognizant of his behaviour, is what causes the observer to feel shame on his behalf. He thus feels shame based on what the foolish man would feel, rather than on what he does feel. Again, we see that imaginative sympathy is necessary in this case because it enables the sympathizer to make the requisite causal inference that gives rise to the initial idea of the other’s (hypothetical) passion. As with the previous example, if we sympathized using a contagion-like process, sympathy would have to be with what the person was actually feeling, not what we think he would be feeling. We could not feel shame as a result because the foolish person does not himself feel shame. Although this is an example in which the sympathizer feels humility (shame) on behalf of the other person, it appears during Hume’s discussion of compassion as an example of sympathizing with ‘affections which have no existence’. (Treatise, p. 370)

2.3.2.3 Example 3

In Example 1, the man’s greatness of mind is what prevented him from feeling sorrow on account of his misfortune. We sympathize with the passion that misfortune usually engenders, the man’s greatness of mind
notwithstanding. In Example 2, we sympathize with the passion the foolish man would feel if he were aware of his behaviour. In this example, Example 3, we sympathize with feelings that do not exist at all. The sleeper who is murdered and the infant prince who is too young to know that he is being held captive by his enemies are unaware of their respective situations, and so are not feeling any emotion as a result. The sympathizer infers sorrow (effect) from the situation (cause).

‘Tis an aggravation of a murder, that it was committed upon persons asleep and in perfect security; as historians readily observe of any infant prince, who is captive in the hands of his enemies, that he is more worthy of compassion the less sensible he is of his miserable condition. As we ourselves are here acquainted with the wretched situation of the person, it gives us a lively idea and sensation of sorrow, which is the passion that generally attends it; and this idea becomes still more lively, and the sensation more violent by contrast with that security and indifference, which we observe in the person himself. A contrast of any kind never fails to affect the imagination, especially when presented by the subject; and ‘tis on the imagination that pity entirely depends. (Treatise, p. 371, Hume’s italics)

In this quotation we can see firstly that Hume himself mentions compassion’s reliance on the imagination, and secondly that LLE could not give rise to compassion. This is because the sympathizer’s compassion arises as a result of an initial inference from the situation of the victim to the usual effect, an inference which requires using the imagination.

2.3.2.4 Example 4

In this example, the object of the sympathy is also unaware of his condition—imminent danger—because he is asleep. The person’s misfortune is anticipated, not actual. This case is also one of inferring from the future cause (being trampled by horses) to its future effect (for example, pain or sorrow), and converting this into our own impression of pain. We
have deemed the sleeper’s situation to be dangerous, and the idea of his imminent danger has caused us to sympathize.

For supposing I saw a person perfectly unknown to me, who, while asleep in the fields, was in danger of being trod under foot by horses, I shou’d immediately run to his assistance; and in this I shou’d be actuated by the same principle of sympathy, which makes me concern’d for the present sorrows of a stranger. The bare mention of this is sufficient. (*Treatise*, p. 385)

2.3.2.5 Example 5

The final example contains two instances in which we can sympathize with an entire group of people and come to feel benevolence for them. In the first case, we sympathize with the residents of a barren and desolate country, inferring from this desolation and barrenness (cause) to their moderate or weak misery or uneasiness (effect). In the second case we infer from the cause (razed city) to the usual effect on the inhabitants (great misery).

A barren or desolate country always seems ugly and disagreeable, and commonly inspires us with contempt for the inhabitants. This deformity, however, proceeds in a great measure from a sympathy with the inhabitants [...] but it is only a weak one, and reaches no farther than the immediate sensation, which is disagreeable. The view of a city in ashes conveys benevolent sentiments; because we there enter so deep into the interests of the miserable inhabitants, as to wish for their prosperity, as well as feel their adversity. (*Treatise*, p. 388)

What these examples show is that imaginative sympathy is necessary for compassion, and for benevolence when this is for strangers or those to whom we do not already bear some connection. In the next section, I argue that imaginative sympathy cannot ensure that compassion is produced in response to another’s misery or suffering because what we feel as a result of

---

50 See *Figure 5* and the surrounding discussion for an explanation of how benevolence is related to love.
it depends on the disposition and character—the motivations—of the sympathizer.

2.3.3 Sympathy, Compassion and Motive

Hume argues that sympathy is sufficient for compassion. Simply sympathizing can cause us to feel compassion on account of another’s misery by converting the idea of another’s misery into extensive sympathy, which is related to compassion. In this section I apply aspect (2) of the motivation objection: I argue that a person’s disposition and character determines whether or not he or she can feel the extensive sympathy that leads to compassion, which means that sympathy is not, in actual fact, sufficient for compassion.

Sympathy is the conversion of the initial idea of another’s passion into that passion in ourselves. To feel compassion, the initial idea must be such that it causes extensive sympathy. In the case of poverty, for example, if we are unable to ‘extend’ our sympathy, we will feel contempt for another’s poverty rather than compassion or benevolence. Obtaining the initial idea depends on making an inference from cause to effect or effect to cause. These causal inferences are based on custom and habit in the form of general rules. However, it transpires that a sympathizer’s ability to make the right initial causal inference depends on his disposition and character, for it is these that determine our ability to be influenced by the general rules that tell us which causes ought to go with which effects. In the section entitled ‘Limitations of this system’ (Treatise, pp. 290-294), Hume explains the influence of general rules on the passions.

I may add as a fifth limitation, or rather enlargement of this system, that general rules have a great influence upon pride and humility, as well as on all the other passions. Hence we form a notion of different ranks of men, suitable to the power or riches
they are possess of; and this notion we change not upon account of any peculiarities of the health or temper of the persons, which may deprive them of all enjoyment in their possessions. This may be accounted for from the same principles, that explain’d the influence of general rules on the understanding. Custom readily carries us beyond the just bounds in our passions, as well as in our reasonings.

(Treatise, p. 293, Hume’s italics)

Hume discusses causation in Book I of the Treatise. Only causation can ‘give us assurance from the existence or action of one object, that ‘twas follow’d or preceded by any other existence or action.’ (Treatise, p. 74) It is ‘the only [relation], that can be trac’d beyond our senses, and informs us of existences and objects, which we do not see or feel.’ (Treatise, p. 74) When we make an inference from the observation of a cause to its effect or vice versa, we do so based on our experience of having observed their constant conjunction in the past: ‘we call the one cause and the other effect, and infer the existence of the one from that of the other. In all those instances, from which we learn the conjunction of particular causes and effects, both the causes and effects have been perceiv’d by the senses, and are remember’d: But in all cases, wherein we reason concerning them, there is only one perceiv’d or remember’d, and the other is supply’d in conformity to our past experience.’ (Treatise, pp. 87, 90)

It is because of the principles of association (contiguity, causation and resemblance) operating on us, rather than by reasoning, that our mind passes from a cause to its usual effect or vice versa. (Treatise, p. 92) The principles of association ensure that ‘when ev’ry individual of any species of objects is found by experience to be constantly united with an individual of another species, the appearance of any new individual of either species naturally conveys the thought to its usual attendant.’ (Treatise, p. 93) This association between ideas is based on custom; our minds automatically pass from the
appearance of one idea or object to the existence of the one that our minds
are accustomed to it appearing with. (*Treatise*, p. 103)

Custom influences our judgements concerning cause and effect
automatically and pre-reflectively (*Treatise*, pp. 93, 104, 147)\(^{51}\) and, in the
form of general rules, influences *all* our causal inferences. We can be
influenced by general rules even when they are ‘contrary to present
observation and experience.’ (*Treatise*, p. 147) It can also be the case that
custom produces a ‘contrariety in our sentiments concerning the same
object.’ (*Treatise*, p. 147)\(^{52}\) Thus a man who is suspended in an iron cage can
believe he is going to fall to his death and simultaneously believe that he is
securely tied and therefore not in any actual danger of falling.\(^{53}\) Hume
argues that this contrariety arises as a result of our being (pre-reflectively)
affected by custom in the form of *two sets* of general rules. The imagination
can run away with itself, so to speak, and be governed by a false set of
general rules without our noticing, thus causing conflict with those general
rules ‘by which we ought to regulate our judgement concerning causes and

\(^{51}\) The custom operates before we have time for reflexion. The objects seem so inseparable, that
we interpose not a moment’s delay in passing from the one to the other. But as this transition
proceeds from experience, and not from any primary connexion betwixt ideas, we must necessarily
acknowledge, that experience may produce a belief and a judgement of causes and effects by a
secret operation, and without being once thought of.’ (*Treatise*, p. 104)

\(^{52}\) ‘Our judgements concerning cause and effect are deriv’d from habit and experience; and when we
have been accustomed to see one object united to another, our imagination passes from the first to
the second, by a natural transition, which precedes reflection, and which cannot be prevented by it.
Now ‘tis the nature of custom not only to operate with its full force, when objects are presented,
that are exactly the same with those to which we have been accustom’d; but also to operate in an
inferior degree, when we discover such as are similar; and tho’ the habit loses somewhat of its force
by every difference, yet ‘tis seldom entirely destroy’d, where any considerable circumstances remain
the same [...] From this principle I have accounted for that species of probability, deriv’d from
analogy, where we transfer our experience in past instances to objects which are resembling, but
are not exactly the same with those concerning which we have had experience. In proportion as the
resemblance decays, the probability diminishes; but still has some force as long as there remain any
traces of resemblance. This observation we may carry father; and may remark, that tho’ custom be
the foundation of all our judgements, yet sometimes it has the effect on the imagination in
opposition to the judgement, and produces a contrariety in our sentiments concerning the same
object.’ (*Treatise*, p. 147)

\(^{53}\) Beliefs are, on Hume’s view, sentiments or impressions that arise from causation. See e.g.
*Treatise* (p. 107).
effects.’ (Treatise, p. 149, my italics) How susceptible we are to the influence of the ‘wrong’ general rules depends on our disposition and character. The vulgar are more influenced by the spurious general rules, whereas the wise are more easily influenced by the proper ones.

Thus our general rules are in a manner set in opposition to each other. When an object appears, that resembles any cause in very considerable circumstances, the imagination naturally carries us to a lively conception of the usual effect, tho’ the object be different in the most material and most efficacious circumstances from that cause. Here is the first influence of general rules. But when we take a review of this act of the mind, and compare it with the more general and authentic operations of the understanding, we find it to be of an irregular nature, and destructive of all the most establish’d principles of reasoning; which is the cause of our rejecting it. This is a second influence of general rules, and implies the condemnation of the former. Sometimes the one, sometimes the other prevails, according to the disposition and character of the person. The vulgar are commonly guided by the first, and wise men by the second. (Treatise, pp. 149-150, my italics)

According to Hume, as we have seen, the reason sympathizers obtain a weak idea of another’s misery is down to the fact that ‘a certain degree of poverty produces contempt; but a degree beyond causes pity and good-will.’ (Treatise, p. 387) As a result, ‘when the misery of a beggar appears very great, or is painted in very lively colours, we sympathize with him in his afflictions, and feel in our heart evident touches of pity and benevolence.’ (Treatise, p. 387) The same object, Hume claims, ‘causes contrary passions according to its different degrees.’ (Ibid.) When we feel contempt, we infer from the cause (slight poverty) to its effect (weak misery) and get a weak idea, which converts into limited sympathy. In compassion, we infer from the cause (great poverty) to its effect (great misery) and get a strong idea, which converts into extensive sympathy.
It is, given the fact that custom can automatically and pre-reflexively produce a ‘contrariety in our sentiments concerning the same object’, reasonable to suggest that someone in slight poverty could cause us to feel compassion when what we ought to feel—according to Hume’s system—is contempt. Because of our disposition and character, we could fail to be influenced by the correct set of general rules that would enable us to infer from the beggar’s degree of poverty to its usual effect, the level of misery that usually attends it. We would therefore end up feeling the wrong sentiment as a result of our sympathy.

This is supported by what Hume says during his discussion of compassion. There, he claims that other facts about us, such as our gender and age, determine how likely we are to be influenced by the wrong general rules. It is because women and children are more susceptible to the influence of the wrong general rules that they ‘pity extremely those, whom they find in any grief or affliction’ (Treatise, p. 370) and ‘faint at the sight of a naked sword, tho’ in the hands of their best friend.’ (Ibid.) So they could feel compassion for a beggar in slight poverty when, according to Hume’s system, they should end up feeling contempt.

It would not, one may well argue, be such a bad thing to fail to feel contempt, and instead to feel compassion. But it is reasonable to suggest, given Hume’s system, that there is a possibility of our feeling contempt for a beggar in severe poverty when we ought to feel compassion, should we be disposed to be influenced by the wrong general rules. And this outcome would be much less palatable. For example, perhaps vulgar, unwise men are more likely to feel contempt because they are more easily influenced by the wrong general rules, and will thus fail to infer from the cause (the abject poverty of the beggar) to its usual effect (great misery). The conclusion to
draw from this is that our disposition and character mean that sympathizing cannot guarantee that we feel compassion for another’s misery; sometimes it can produce contempt, if our disposition and character mean we are susceptible to being influenced by the wrong general rules.

2.3.4 Summary

In this section I provided exegesis and analysis of sympathy’s role in the production of compassion and benevolence. I introduced extensive and limited sympathy and put forward an interpretation of their production that did not involve an extra ‘first sympathy’. I argued that Hume does have an account of how sympathy produces concern, but that in order to accept it we must also accept his associationism. I also argued that imaginative sympathy is necessary for compassion. Finally, I argued that this form of sympathy is not sufficient for compassion, as Hume claims, because it cannot guarantee that we feel compassion: the disposition and character of the sympathizer can affect the outcome of the sympathizing. Those with a certain character and disposition can fail to make the correct initial inference required for feeling the correct passion because they are more easily influenced by the wrong set of general rules. The result is, for example, that we can sometimes fail to feel compassion as a result of sympathy when, according to Hume’s system, we should do so. Thus sympathy cannot be sufficient for compassion.

2.4 Moral Evaluation

In this section I apply aspect (2) of the motivation objection to Hume’s theory of moral judgement. In the previous section, the claim was that sympathy is insufficient for compassion. Here, however, the claim is that sympathy is not sufficient for moral judgement. I argue that the sympathizer’s motives—his disposition and character—mean that his sympathizing cannot guarantee that he makes a moral judgement. On Hume’s view, making a moral
judgement requires feeling a moral sentiment, which we can only feel if we adopt a common or general point of view. Only those with a certain disposition and character are able to discern moral sentiments, felt from the common point of view, from non-moral sentiments, felt from their own point of view. As a result, many people could fail to make moral judgements because they fail to discern the moral sentiments from their own sentiments. Hume could, however, accept this. But he faces a more fundamental problem: why would a person who can make moral judgements on a reliable basis ever elect to privilege his moral sentiments over his own, non-moral sentiments? We lack an account of why, on Hume’s view, anyone would ever be motivated to be moral, or have a reason to act morally, which looks problematic for his claim that moral judgements can motivate us to be moral. For if it is the case that we can make evaluations from our own perspectives—say we approve of immoral character traits—why would we not use this as a guide to our moral action, or cultivate these non-moral character traits, rather than the character traits we know to be moral because we feel the moral sentiment towards them? I consider a possible way out of this quandary for Hume, viz. the positing of a motive of universal benevolence that would provide sympathizers with the motivation to privilege the moral sentiments. However, I conclude that this possibility is not open to him, because he denies that such a motive exists independently of what generates the moral sentiments in the first place: sympathy. In this section I also defend the view that Hume is consistent in his use of sympathy in Books II and III of the Treatise.

2.4.1 Sympathy and the Moral Sentiments

In this section I explain sympathy’s role in the production of the moral sentiments. On Hume’s view, approving and disapproving of other people’s actions and character traits, and thereby denominating them virtuous or vicious, is a matter of feeling a ‘peculiar’ kind of sentiment, a moral
sentiment of approbation or disapprobation, towards them. (*Treatise*, p. 472)
We call someone’s action or character trait vicious or virtuous if we feel this special sentiment towards it. We do not infer from the fact that we feel this way towards an action or character trait that it is a vice or a virtue. Rather, the feeling is itself our approval of the trait as a virtue, or our disapproval of it as a vice. Hume explains this as follows.

An action, or sentiment, or character is virtuous or vicious; why? because its view causes a pleasure or uneasiness of a particular kind. In giving a reason, therefore, for the pleasure or uneasiness, we sufficiently explain the vice or virtue. To have the sense of virtue, is nothing but to feel a satisfaction of a particular kind from the contemplation of a character. The very feeling constitutes our praise or admiration. We go no farther; nor do we enquire into the cause of the satisfaction, we do not infer a character to be virtuous, because it pleases: But in feeling that it pleases after such a particular manner, we in effect feel that it is virtuous. The case is the same as in our judgements concerning all kinds of beauty, and tastes, and sensations. Our approbation is imply’d in the immediate pleasure they convey to us. (*Treatise*, p. 471, Hume’s italics)

Our moral sentiments are peculiar feelings of pleasure and pain. (*Treatise*, p. 472) Moral sentiments are a particular type of sentiment, the token instances of which can differ according to the object that gives rise to them. They all resemble each other sufficiently to be accorded the title of ‘moral’ sentiment, however. (*Treatise*, p. 617) They arise from the disinterested contemplation of a person’s character traits, or of his actions, as the outward signs of the possession of those traits. (*Treatise*, pp. 477, 575) ‘Disinterested’ contemplation is contemplation that is ‘without reference to our own particular interest.’ (*Ibid.*) Only those sentiments that are felt independently of our own interest count as moral sentiments. (*Treatise*, p. 472)

Nor is every sentiment of pleasure or pain, which arises from characters and actions, of that peculiar kind, which makes us praise or condemn. The good qualities of an
enemy are hurtful to us; but may still command our esteem and respect. 'Tis only when a character is considered in general, without reference to our particular interest, that it causes such a feeling or sentiment, as denominates it morally good or evil. (*Treatise*, p. 472, Hume’s italics)

A discerning judge can tell the difference between feelings of pleasure or pain that arise when we contemplate another’s character with our own interests in mind, or those of our nearest and dearest, and those that arise because of the contemplation of his traits independently of such interests. (*Treatise*, p. 472)

The moral sentiments always give rise to one of the indirect passions of pride, humility, love and hatred, because the same conditions obtain for the production of vice and virtue as for the production of these passions. (*Treatise*, p. 473) In other words, the conditions are right for the operation of the double relation of ideas and impressions, explained in §2.2.1.2. When we feel the moral sentiments, we feel them when we disinterestedly contemplate an action or trait which bears some relation either to ourselves (if it is our own trait, for example) or another person, and produces a sensation that is related to the characteristic sensation of the indirect passion. (*Treatise*, p. 473) The moral sentiments of approval arise, according to Hume, when a quality of the mind (durable character trait) is: (1) immediately agreeable to others; (2) immediately agreeable to its possessor; (3) useful to its possessor, or (4) useful to others. (*Treatise*, p. 591) *Mutatis mutandis* for disapproval. The moral sentiment of approval ‘may arise either from the mere species or appearance of characters and passions, or from reflexions on their tendency to the happiness of mankind, and of particular persons.’

---

54 This will be of particular relevance in §2.4.3.
55 There is a debate surrounding whether the moral sentiments give rise to the indirect passions, or whether they are themselves versions of these passions. I take the former view here, based on evidence to be found in the *Treatise* (pp. 473, 575, 589, 591). See Loeb (1977) for an introduction to this debate.
(Treatise, p. 589) I will now briefly outline how the moral sentiments arise in each of these ways.

2.4.1.1 Traits immediately agreeable to others

Some character traits give rise to our moral approval (or disapproval) by ‘original principles of human nature, which cannot be accounted for.’ (Treatise, p. 590) Of this kind is, for example, wit. A person’s wit makes him immediately agreeable—and therefore causes pleasure in—those around him, and this pleasure constitutes their approval of him. (Treatise, p. 590)

Sympathy plays a role in producing feelings of approval (and disapproval) when we approve of people because their traits are agreeable to those who come into contact with them, even on occasions when we ourselves have not ‘reap’d any pleasure from them.’ (Treatise, p. 590) To feel the moral sentiment of approval for another person because it is immediately agreeable, we must feel pleasure because his trait causes it in us. Sympathy enables A to feel approval of B in virtue of the pleasure his trait causes to others, even though he himself has not felt pleasure, because he is not in the witty person’s company, for example. This occurs via the sympathy mechanism: A infers from B’s wit (cause) to its usual effect on those who would come into contact with him, which would be pleasure. The sympathy mechanism enables A to convert this into his own pleasure, which constitutes his approval of B’s trait.

2.4.1.2 Traits immediately agreeable to their possessor

Hume claims that ‘each of the passions and operations of the mind has a particular feeling’ which is pleasant or painful. (Treatise, p. 590) So if a person is witty, then, necessarily, this passion will feel pleasurable to its possessor, and others will approve of the possessor because he possesses a trait that
causes him pleasure, a pleasure they feel via sympathy. *(Treatise, p. 590)*

This pleasure constitutes their approval.

Sympathy also enables us to approve of a person for traits that are immediately agreeable to him, but are ‘of no service to any mortal’. *(Treatise, p. 590)* For example, A can approve of B’s wit, a trait that *would have been* agreeable to him, even though he is no longer living, say, and cannot therefore derive any pleasure from it. In this case, A infers from the actual or usual effect of B’s trait on B himself, pleasure, and converts it into pleasure, which constitutes his approval of B’s trait.

2.4.1.3 Traits useful to their possessor

Hume claims that, if we were to take a survey of the qualities we usually attribute to those whom we admire, we would discover that they normally fall into two categories: traits that ‘make [people] perform their part in society’ on the one hand, and traits that ‘render them serviceable to themselves, and enable them to promote their own interest,’ on the other. *(Treatise, p. 587)* Of the first kind are generosity and humanity, which I consider below (in §2.4.1.4). Of the second are, among others, prudence, temperance, frugality, and industry. *(Treatise, p. 587)* When B has a great ‘dexterity in business’, a trait that is useful to him, A approves of this trait because it is a means to an agreeable end, that end being the happiness and good of B. *(Treatise, p. 588)* The end—B’s pleasure or good—must therefore be agreeable and important to A. A does not know B, so he does not care about his good, or take an interest in him, simply because he loves him.

---

56 This raises the following problem in relation to compassion. If compassion, as we saw earlier, has a painful feeling, then do we disapprove of others for possessing the virtue of compassion, because it must necessarily cause them pain? Perhaps Hume would respond that this is a trait that we would approve of because it is useful to others, in which case its phenomenology would not matter.
Rather, he comes to care about B’s good by sympathizing with him. As Hume puts it:

The person is a stranger: I am in no way interested in him, nor lie under any obligation to him: His happiness concerns not me, farther than the happiness of every human, and indeed of every sensible creature: That is, it affects me only by sympathy. From that principle, whenever I discover his happiness and good, whether in its causes or effects, I enter so deeply into it, that it gives me a sensible emotion. The appearance of qualities, that have a *tendency* to promote it, have an agreeable effect upon my imagination, and command my love and esteem. (*Treatise*, pp. 588-9, Hume’s italics)

When approving of B’s dexterity in business, A infers from the cause (dexterity) to its usual effect (pleasure), which gives him the ‘sensible emotion’ of pleasure, which constitutes his moral sentiment of approval. This then gives rise to love and esteem.

2.4.1.4 Traits useful to others

Sympathy also accounts for why we approve of the so-called ‘social virtues’, those virtues which are generally deemed to be good for society. These include: beneficence, charity, generosity, clemency, moderation and equity. (*Treatise*, p. 578) According to Hume, we approve of these virtues because they tend to the interest (good) of society, which we care about because of sympathy.

Moral distinctions arise, in a great measure, from the tendency of qualities and character traits to the interest of society, and [...] ’tis our concern for that interest, which makes us approve or disapprove of them. Now we have no such *extensive* concern for society but from sympathy; and consequently ’tis that principle, which takes us so far out of ourselves, as to give us the same pleasure or uneasiness in the

---

57 I have been dealing with Hume’s natural virtues in this chapter. Sympathy is also responsible for why we approve of justice, an artificial virtue. We approve of just acts because we consider the institution of justice to be beneficial to society. See especially *Treatise* (pp. 499, 579).
characters of others, as if they had a tendency to our own advantage and loss. 

(Treatise, p. 579, Hume’s italics)

If, for example, B acts generously towards C, A will approve of B for his generosity because generosity is a trait that tends to benefit society. A infers from the cause of B’s action (his generous character) to its usual effect (happiness/good/pleasure of society). This gives him an idea of the potential pleasure (happiness) of society, which is vivified by his idea of self and the principles of association, such that he converts it into pleasure, which is the moral sentiment of approval. If we contemplate another’s trait and infer that it has harmful effects on its possessor or on society, then we feel disapproval because we convert an idea of the other’s potential pain into moral disapproval. Figure 7 below depicts the connection between the moral sentiment of approval and the indirect passion of love.
Figure 7: Extensive Sympathy and Moral Approval

KEY:

********** = Related by resemblance

2.4.1.5 The Abandonment Thesis

Before introducing the notion of the common point of view, and arguing that Hume faces the motivation objection in relation to his theory of moral approval, I briefly defend the claim that Hume is consistent in his use of sympathy throughout Books II and III of the Treatise.

Not all commentators agree that Hume is consistent in his use of sympathy throughout the Treatise. Proponents of the so-called ‘abandonment thesis’ argue that Hume actually jettisons the sympathy of Book II—limited sympathy—for the extensive sympathy that appears in Book III to ground morality. Various commentators, for example Philip Mercer (1972), John
Bricke (1997) and Jennifer Herdt (1997), have argued that certain features of Book II’s limited sympathy render it deeply unsatisfactory. They argue, for example, that limited sympathy only enables sympathy with another person’s ‘current mental state, detached from his or her diachronic psychological life’ (Cunningham, 2004, p. 237) and as such cannot include the kind of concern for another person or for society in general that Hume claims is fundamental for morality. The abandonment thesis is put forward as a solution to what Andrew Cunningham (2004, p. 238) has described as the ‘weak sympathy paradox’, which asks why Hume would have given ‘such potency to a form of sympathy that he recognized as limited.’ Abandonment theorists argue that it is because he recognised the limitations and weaknesses of limited sympathy that Hume drops it in favour of extensive sympathy in Book III.

In my view, Hume does not abandon limited sympathy for extensive sympathy. As we saw earlier, in §2.3.1.1, limited and extensive sympathy both exist in Book II and are, as I have already argued (in §2.3.1.1), two different outputs of the sympathy mechanism. Cunningham rightly rejects the abandonment thesis. However, I disagree with his principal reason for doing so. He takes the view that there are three forms of sympathy: one limited and two extensive. Hume, Cunningham argues, uses limited and extensive in Book II, and then a third form of sympathy in Book III. This third form is still consistent with Book II’s sympathy because it is simply another form of extensive sympathy, something Hume has already outlined in Book II. Cunningham argues that there is a ‘crucial distinction between the two types of extensive sympathy,’ which is that they ‘extend’ over different things. (Cunningham, 2004, p. 243) Book II’s extensive sympathy ‘refers to sympathy extending along the temporal dimension of a single human being—sympathy that extends beyond the communication of a
passing mental state so as to produce an acquaintance with, or understanding of, the diachronic person.’ (Ibid.) Book III’s extensive sympathy, on the other hand, ‘refers to sympathy that potentially extends to any or all members of society.’ (Ibid.)

I agree that extensive sympathy can range over individuals and society. However, I disagree with Cunningham that this means there are two forms of extensive sympathy. The extensive sympathy that ranges over society does not suddenly appear in Book III, after having been absent from Book II. For, as we saw in §2.3.2, Book II’s extensive sympathy is capable of ranging over both individuals and society. I therefore disagree that Book III’s extensive sympathy is a third form of sympathy. Book III’s sympathy, the sympathy at work in moral approval, is the sympathy mechanism, introduced in Book II, and the same extensive sympathy from Book II. It is not another form of extensive sympathy. Hume uses ‘sympathy’ to refer to sympathy as the mechanism that converts an idea into an impression (Treatise, pp. 317, 369) and to the output of this process, for example when he refers to ‘the communicated passion of sympathy.’ (Treatise, p. 370) Perhaps this tendency to use the same term to refer to different phenomena is what gives rise (a) to the abandonment theorists’ claim that Book II sympathy is abandoned in Book III and (b) to Cunningham’s claim that there is another form of extensive sympathy operating in Book III. Book III sympathy is, in my view, the mechanism of Book II, with its outputs, limited sympathy and extensive sympathy.

2.4.2 The Common Point of View

We have seen that approval and disapproval of character traits is a matter of feeling a moral sentiment, which is a certain type of pleasure or pain, when

---

we contemplate those traits. We also saw that, in order to count as moral sentiments, traits must be contemplated independently of our own interest. On Hume’s view, the way we make moral judgements that are disinterested is by making them from a so-called common or general point of view.\footnote{Hume refers to the common point of view as ‘steady’ and ‘general’. See e.g. \textit{Treatise}, pp. 581, 583 I will be referring to it as the common point of view in what follows.}

Hume introduces the notion of the common point of view in Book III of the \textit{Treatise}. It is introduced to deal with ‘two remarkable circumstances […] which may seem objections’ to his account of sympathy’s role in the production of our moral sentiments. \textit{(Treatise, p. 580)} I will call these objections ‘virtue in rags’ and ‘variability’.\footnote{Here I am following Cohon (2008b, p. 131).}

\subsection*{2.4.2.1 Variability}

The variability objection arises because, although the outputs of our sympathy are variable, our moral judgements are stable. \textit{(Treatise, p. 581)} Sympathy is governed by the relations, as we saw. This means, for example, that the stronger the relations of contiguity and resemblance are between A and B, the easier it will be for A to sympathize with B. If sympathy really did give rise to the moral sentiments, as Hume claims, then we must surely be more likely to approve of, and consider more virtuous, those we bear closer relations to than those to whom we bear no connection at all, simply because we will love them already: they are our nearest and dearest. \textit{(Treatise, pp. 317-318)} However, Hume observes that: ‘we give the same approbation to the same moral qualities in China as in England. They appear equally virtuous, and recommend themselves equally to the esteem of a judicious spectator.’ \textit{(Treatise, p. 581)} So it would appear that our ‘sympathy varies without a variation in our esteem.’ \textit{(Treatise, p. 581)} As Hume puts it:
'tho sympathy be much fainter than our concern for ourselves, and a sympathy with persons remote from us much fainter than that with persons near and contiguous; yet we neglect all these differences in our calm judgements concerning the characters of men. (Treatise, p. 603)

Furthermore, we tend to love our friends and relations on account of their virtues more than historical figures who are famous for their virtue(s). Yet we still approve of the actual virtue in historical figures as much as we do the same virtue in those closer to us in time. There is thus a variation in our sympathizing but no corresponding variation in our approval. The outputs of our sympathy are variable, but our moral judgements tend to be steady and consistent. And if this is true, then it cannot also be true, as Hume claims, that sympathy is the chief source of moral distinctions. (Treatise, p. 579)

2.4.2.2 Virtue in Rags

The second problem relates to Hume’s claim that to approve of a virtue because it benefits society we have to see it manifested in action, in order to be able to infer from the cause (the trait) to its usual effect (happiness). If we did not get the opportunity to see these traits manifested in actions, we could not approve of a person for possessing them, and could not therefore denominate him virtuous for possessing the trait. However, sometimes a person’s circumstances prevent him from ever exercising his virtuous traits. If an individual is poor, for example, he cannot be generous with his money, even if he wants to be. Nevertheless, Hume claims, we still call a person virtuous and approve of and esteem him for possessing the virtue even though he has no opportunity to exercise it. (Treatise, p. 584) And if this is true, then it cannot also be true that our approval of virtues that tend to the good of society is down to sympathy.
2.4.2.3 Hume’s Solution

Hume’s solution is to invoke the common point of view. The reason we can make moral judgements that are steady, even though we make them by sympathizing, which is variable because governed by the relations, is that in the case of moral judgement, we are sympathizing not from our own point of view, but rather from a common point of view. When we make our moral judgements, Hume argues, we ‘fix on some steady and general points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation.’ (Treatise, p. 583)

Our situation, with regard both to persons and things, is in continual fluctuation; and a man, that lies at a distance from us, may, in a little time, become a familiar acquaintance. Besides, every particular man has a peculiar position with regard to others; and ‘tis impossible we cou’d ever converse together on reasonable terms, were each of us to consider characters and persons, only as they appear from his peculiar point of view. In order, therefore, to prevent those continual contradictions, and arrive at a more stable judgement of things, we fix on some steady and general points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation. (Treatise, pp. 581-2, Hume’s italics)

Because our situation is in continual fluctuation, it will always be the case that our relationships with others, and our own interests, change. Given that our moral evaluations are based on our sentiments, and our sentiments can vary depending on our current situation, the only way we can make genuine moral evaluations that are constant and can withstand such fluctuations is to adopt the common point of view, and judge others from that vantage point. The common point of view is thus a device for correcting the sentiments we have when we contemplate character traits from our own, ‘situated’ point of view.61 The common point of view in moral judgement operates analogously.

---

61 Here I follow Cohon (2008b, p. 135) who uses the term ‘situated’ to describe the sentiments we feel from our own, ‘interested’ points of view.
to the one we adopt when we make aesthetic judgements. It will therefore be useful to briefly consider the common point of view as it operates in aesthetic judgement.

2.4.2.4 Moral and Aesthetic Judgement

On Hume’s view, aesthetic qualities and moral qualities are both judged according to the pleasure or pain they cause in us when we see them, or reflect on their tendency to produce pleasure or pain in others. *(Treatise*, pp. 364-5, 469, 576, 585, 618) Sentiments of moral approval and disapproval proceed ‘entirely from a moral taste, and from certain sentiments of pleasure or disgust, which arise upon contemplation and view of particular qualities or characters.’ *(Treatise*, p. 581) Similarly, our sense of beauty proceeds from certain sentiments of pleasure or disgust which arise when we see or reflect on the tendencies of certain aesthetic qualities to cause pleasure or pain. We correct our sentiments of beauty to correct for distortions due to our own perspectives. For example, we still say a person is beautiful even though she is at twenty paces, and we cannot derive as much pleasure from seeing her from this distance as we would if she were standing in front of us. *(Treatise*, p. 582) Such correction and adjustment of our situated sentiments is ‘common with regard to all the senses; and indeed ‘twere impossible we could ever make use of language, or communicate our sentiments to one another, did we not correct the momentary appearances of things, and overlook our present situation.’ *(Treatise*, p. 582) We adjust for our situated sentiments of morals as well as for our sentiments of beauty using a common point of view.

2.4.2.5 The ‘Two-Feeling’ Interpretation

We have seen that, in order to explain how it is that we can make stable moral judgements by sympathizing, Hume invokes the device of the common point of view. I interpret Hume as claiming that we sympathize
from this common point of view when we make moral judgements and that we also sympathize from our own, situated perspectives. We therefore feel two sentiments: a situated, non-moral sentiment and a moral sentiment. This is akin to the way in which we make aesthetic judgements: we ‘correct’ our own, situated emotional response with the one we have from the common point of view.

Not all commentators interpret Hume this way, however. Barry Stroud (1977, p. 191), for example, argues that ‘our moral judgements, like our aesthetic judgements, are not always direct expressions of our actual feelings.’ This is because, he claims, we sometimes do not feel anything from the common point of view. Our actual moral judgements can be based on beliefs about what we would feel if we were to occupy this point of view, and did not occupy our own point of view. We learn by induction from past experience how we would feel, or how we should feel, in response to certain character traits, if we were to contemplate these traits from the common point of view. We sometimes fail to feel a sentiment from the common point of view, and instead replace our situated sentiment with an inductively grounded belief, and this constitutes our moral judgement.

Stroud (ibid.) concludes that ‘Hume’s theory has the effect of divorcing our moral judgements from our actual feelings.’ This is a problem for Hume in two ways. Firstly, Hume claims that morality is ‘more properly felt than judg’d of’ (Treatise, p. 470), which implies that we feel a sentiment every time we make a moral judgement. If our moral judgements are actually based on beliefs, not sentiments, then it cannot always be the case that our moral judgements have an affective component. Secondly, Hume claims that moral judgements have the power to motivate us. Reason, he famously argues, has no such effect.
Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason. (Treatise, p. 457)

Given that sentiments alone can motivate action (Treatise, pp. 457, 459) then it must surely be the case that our moral judgements contain feelings on every occasion. If we made moral judgements from the common point of view, from which we can fail to feel a sentiment, according to Stroud, then it cannot be the case that our judgements always motivate us to act. Stroud (1977, p. 192) concludes that Hume ‘shows no real awareness of these difficulties in his theory,’ and does not have the resources to resolve them.

The claim that our moral judgements can be based on beliefs and not sentiments, and in fact often override our situated sentiments, does seem to tally with what Hume says on several occasions in the Treatise about the heart not always following the head. (Cf. Cohon, 2008b, p. 139) These passages seem to support the kind of view Stroud holds. For example, Hume claims that ‘experience soon teaches us this method of correcting our sentiments, or at least, of correcting our language, where the sentiments are more stubborn and inalterable.’ (Treatise, p. 582) If our passions can remain fixed and unchangeable, this implies that we must sometimes supplant them with a moral belief, when the passions remain intractable. Hume also claims that ‘the passions do not always follow our corrections; but these corrections serve sufficiently to regulate our abstract notions, and are alone regarded, when we pronounce in general concerning the degrees of vice and virtue.’ (Treatise, p. 585) Finally, Hume says the following: although ‘the heart does not always take part with those general notions, or regulate its love and hatred by them, yet are they sufficient for discourse, and serve all our purposes in company, in the pulpit, on the theatre, and in the schools.’
These excerpts appear to indicate that we must sometimes fail to feel anything from the common point of view and must therefore sometimes rely on inductively grounded beliefs from past experience to tell us what we would feel. (Cf. Cohon, 2008b, p. 139) So it looks like Hume’s claim that moral evaluation is always a matter of feeling is in jeopardy.

Rachel Cohon (2008b, p. 143) has put forward what she calls a ‘two-feeling interpretation’ of Hume’s theory of moral judgement that she claims can explain how moral judgement can still be a matter of feeling and not reason. According to her two-feeling interpretation, we always feel a sentiment—a moral sentiment—from the common point of view. We feel this in addition to a sentiment from our situated perspective. When our hearts fail to follow our heads, it is not because we fail to feel a sentiment from the common point of view, but rather because our moral sentiment is not strong enough to trump our situated one. (Cohon, 2008b, p. 140) When we contemplate a character trait, she argues, we feel a calmer, weaker passion, which is the moral sentiment, from the common point of view, and a stronger, more violent passion, from our situated perspective. (Ibid., p. 139) When our hearts fail to follow our heads, it is because we fail to adhere to the moral sentiment, and base our moral judgement on that calm feeling. We base it instead on our more violent, situated sentiment. (Ibid., pp. 139-140)

I consider Cohon’s to be the more plausible interpretation for three reasons. Firstly, Hume argues that it is possible to feel a moral sentiment alongside a non-moral, situated sentiment. We can feel these contrary passions without one ‘destroying’ the other. (Treatise, pp. 586-7) This supports Cohon’s claim that we always feel two sentiments, and that in cases of failures of moral judgement, our moral sentiment fails to destroy our situated one. A person can feel sentiments from his situated perspective, which stem from his
interest, alongside his moral sentiments, as is the case in the following example.

[…] the fortifications of a city belonging to an enemy are esteemed beautiful upon account of their strength, tho' we could wish that they were entirely destroyed. The imagination adheres to the general view of things, and distinguishes betwixt the feelings they produce, and those which arise from our particular and momentary situation. (Treatise, p. 586)

We saw earlier (in §2.3.3) that the same object can cause contrary passions. In this example, we feel conflicting sentiments: one from interest (A wants B’s fortifications destroyed because he wants to invade the city) and another, approval, also from the strength of the fortifications, on account of their strength. We approve of the fortifications’ strength because they are useful to their possessor, but we still desire that he does not have such strong fortifications from our own, situated perspective.

Secondly, there is evidence to support Cohon’s interpretation in Book II’s discussion of the passions and the will. There, as Cohon (2008b, p. 142) observes, Hume claims that the same object can cause calm and violent passions, depending on our perspective, and the variation in perspective can cause one to change into the other.

For we may observe, that all depends on the situation of the object, and that a variation in this particular will be able to change the calm and the violent passions into each other […] The same good, when near, will cause a violent passion, which, when remote, produces only a calm one. (Treatise, p. 419)

It is plausible to suggest that, if we were to contemplate the same trait from the common point of view, which is ‘remote’ insofar as it is removed or distant or general, and from our own, which is likely to be closer insofar as it
is ‘interested’, then we could feel a calm sentiment from the former, and a violent one from the latter. (Cf. Cohon, 2008b, p. 142) Moreover, Cohon’s view can be supported by the fact that two passions can ‘mingle and unite’ (Treatise, p. 420) and a calm passion can turn into a violent one (Treatise, p. 438). Sometimes ‘the predominant passion swallows up the inferior, and converts it into itself.’ (Treatise, p. 420) So in some cases the calm passion will prevail, and swallow up the inferior, violent passion, but in other cases, the violent passion will prevail.

Finally, and as Cohon (2008b, p. 142) claims, when Hume talks about our hearts not always following our heads, he does not mean that, in these cases, reason prevails over our passions. Rather, he means that our calm passions do not prevail over our violent ones. (Treatise, p. 418) We often mistake conflicts between calm and violent passions for conflicts between passion and reason, Hume argues, because calm and violent passions feel different, and calm passions are themselves often so imperceptible that they ‘produce little emotion in the mind’ and so are not thought to be sentiments at all. (Treatise, pp. 417, 437-8, 536) Nevertheless, the calm passions ‘often determine the will’ and represent ‘a general calm determination of the passions, founded on some distant view or reflexion.’ (Treatise, p. 417) Violent passions are more apt to determine the will, however, but ‘tis often found, that the calm ones, when corroborated by reflection, and seconded by resolution, are able to control [sic] them in their most furious movements.’ (Treatise, p. 419) Strength of mind is the virtue that ‘implies the prevalence of the calm passions above the violent’. (Treatise, p. 418) But even those with strength of mind sometimes fail to act or judge according to their calm passions. (Treatise, p. 418) In cases where our moral evaluations are wrong, it is not because we have failed to feel a moral sentiment; rather it is because our violent, situated sentiments have prevailed. And in cases where we act
against our own interests, or what we feel to be right when we calmly reflect, our violent, situated sentiments have prompted us to act, as opposed to our calmer sentiments. (Treatise, p. 418)

2.4.3 The Problem of Motive

In the previous section I introduced the common point of view. Hume introduces this device to account for an apparent objection to his view that our moral evaluations of others and their traits tend to be stable, even though our sympathetic reactions can vary. The reason our moral evaluations can be made by sympathizing and still be stable is, Hume argues, because, when we make our moral evaluations, we do so from the common point of view. I defended the two-feeling interpretation put forward by Cohon (2008b) according to which, when we make a moral evaluation, we always feel two sentiments: a situated sentiment from our own point of view and a moral sentiment from the common point of view. This interpretation preserves Hume’s two central theses, viz. that moral evaluation is always a matter of feeling, and that our moral evaluations can motivate us. It also preserves him from any charges of rationalism, because it shows how the evaluation we make from the common point of view is always a matter of sentiment, and is not based on a belief about how we would react if we sympathized from the common point of view.

In this section I apply aspect (2) of the motivation objection to Hume’s theory of moral evaluation. I argue that the sympathizer’s motives—his disposition and character—mean that his sympathizing cannot guarantee that he makes a moral judgement. Only those with a certain disposition and character are able to discern moral sentiments, felt from the common point of view, from non-moral sentiments, felt from their own point of view. As a result, many people could fail to make moral judgements because they fail to discern the
moral sentiments from their own sentiments. I will argue that Hume could, however, accept this. But he faces a more fundamental problem: why would a person who can make moral judgements on a reliable basis ever elect to privilege his moral sentiments over his own, non-moral sentiments? Put simply, what motivation would he have to be moral?

We saw that only the sentiments arising from the disinterested contemplation of another person’s character can give rise to the moral sentiments. (Treatise, p. 472) Sympathizing from the common point of view enables us to feel the moral sentiments because it is a ‘general inalterable standard, by which we may approve or disapprove of characters and manners.’ (Treatise, p. 603) Sympathizing from it allows us to abstract away from any consideration of our own interest. However, we also sympathize from our own, situated perspective. Earlier (§2.2.4.5), we saw that the moral sentiments and our situated sentiments are ‘apt to be confounded, and naturally run into one another.’ (Treatise, p. 472) In the following quotation, Hume claims that our disposition and character determine whether or not we can distinguish between the sentiments from the common point of view and those from our own situated perspective.

‘Tis true, those sentiments, from interest and from morals, are apt to be confounded, and naturally run into one another. It seldom happens, that we do not think an enemy vicious, and can distinguish betwixt his opposition to our interest and real villainy or baseness. But this hinders not, but that the sentiments are, in themselves, distinct; and a man of temper and judgement may preserve himself from these illusions. In like manner tho’ ‘tis certain a musical voice is nothing but one that naturally gives a particular kind of pleasure; yet ‘tis difficult for a man to be sensible, that the voice of an enemy is agreeable, or to allow it to be musical. But a person of a fine ear, who has the command of himself, can separate these feelings, and give praise to what deserves it. (Treatise, p. 472, Hume’s italics)
Hume is claiming in this quotation that those of a certain ‘temper and judgement’ can more easily distinguish between the moral sentiment and the situated sentiment. In the case of compassion (see §2.3.3), a person’s motives—his disposition and character— influence how easily he is able to discern between the right and wrong sets of general rules, which affects the outcome of his sympathizing. In the case of moral evaluation, a person’s motives can influence his moral judgements, insofar as they affect how easily he is able to distinguish between the moral and the interested sentiment. Those with self-command (Treatise, p. 472) and strength of mind (Treatise, p. 418), it is reasonable to suggest, will be better able not only to make accurate moral appraisals, but also to act according to those appraisals. Hume claims that ‘morals excite passions, and produce or prevent actions.’ (Treatise, p. 457) So not only will those of a certain character and disposition be more able to ascertain what the right thing to do is, they will also be more likely to act morally, given that their strength of mind will be better able to allow their calm, moral sentiment to prevail over their situated, violent sentiment.

Hume could simply accept this, however. Perhaps those of a certain disposition and character can more easily distinguish the moral from the non-moral sentiments, and perhaps such people will also be more likely to act morally. Nevertheless, a further, more fundamental problem remains. Given that a good judge ends up with two sentiments, one from his situated perspective and one from the common point of view, why would he privilege the one felt from the common point of view over the one felt from his own perspective, and why would he base his moral evaluations and moral actions on that sentiment, and not on his situated sentiment? What possible motivation would he have for doing this?
Commentators have provided various possible alternatives as to why we might be inclined to privilege the sentiments we feel from the common point of view. For example, Geoffrey Sayre-McCord (1994)\(^{62}\) has argued that we would form a social convention to do so, in order to be able to participate fully in genuine moral discourse. For, as Hume claims, because ‘our situation, with regard both to persons and things, is in continual fluctuation’, if we did not find some common point of view from which to make moral appraisals, ‘’tis impossible we cou’d ever converse together on reasonable terms.’ (Treatise, pp. 581-2)

However, this does not address the question of why we should privilege the moral sentiments, and allow these to guide our action. As we saw in §2.4.1, Hume considers people’s character traits to be the main objects of moral appraisal; we approve of the actions of others as outward manifestations of those traits. What reason would a person have, on Hume’s view, to cultivate the virtues, and act virtuously as a result? The discerning judge knows virtue, Hume claims, by feeling moral sentiments when he contemplates those traits in others. But what would make it the case that he was motivated to cultivate those virtues in himself, or to teach them to his children? (Cf. Cohon, 2008b, p. 156) Cohon (2008b) suggests that we might care about, and therefore desire to, cultivate and promote the moral virtues because approving of them in ourselves and others gives us pleasure. So we ‘care’ about the virtues in the sense that they make us feel pleasure. (Ibid., p. 157) If we cultivated them in ourselves, and promoted them in others, then we would feel more pleasure, garnered from more opportunities to feel the pleasurable feeling of approval. The motive to be moral would, therefore, be the pursuit of pleasure.

Feeling pleasure and seeking pleasure are, on Hume’s view, the primary motivators. (Treatise, p. 619) I feel pleasure in response to observing some trait in others, and to possessing it myself, so I have a motive—in the sense of being caused to act—to cultivate only those qualities that cause me most pleasure. However, it is conceivable that even a discerning judge might feel more pleasure upon contemplation of some other set of qualities from his own perspective than those he contemplates from the common point of view. He may, for example, prize and value, and be more likely to feel approval in relation to, other traits because he values them more highly. What motive would he then have for cultivating and following the moral sentiments, and not his situated sentiments?

Hume might respond here that the discerning judge will always be motivated to override his own, situated sentiments, and pursue and cultivate the sentiments felt from the common point of view, even though they cause him less pleasure than his situated sentiments, because he is independently motivated to be moral. That is, he privileges the moral sentiments and acts according to their precepts because he also desires to act morally. However, this option is not open to Hume for two related reasons. Firstly, as will be recalled from §2.3.1, Hume denies the existence of an independent, universal, benevolent motivation or love of mankind that might constitute such a moral motive. He claims that ‘there is no such passion in human minds, as the love of mankind, merely as such, independent of personal qualities, of services, or of relation to ourself.’ (Treatise, p. 481) Secondly, we are only affected by the happiness or misery of others by sympathizing. Only passions can motivate action, so sympathetic conversion of another person’s passion into an impression in ourselves is necessary for us to be moved by that passion. We can see that there is no benevolent motivation without sympathy from the following quotation.
‘Tis true, there is no human, and indeed no sensible, creature, whose happiness or misery does not, in some measure, affect us, when brought near to us, and represented in lively colours: But this proceeds merely from sympathy, and is no proof of such an [sic] universal affection to mankind. (Treatise, p. 481)

So we still lack a reason why the good judge would be motivated to act according to his moral sentiments, given that there is no independent motivation to be moral. Moreover, it looks as though, on Hume’s view, knowing virtue and being motivated to act for the good of others—being benevolent—both depend on the outcomes of the sympathy mechanism. And we have seen that the outcomes of sympathizing depend very much on the character and disposition of the sympathizer. If sympathy is necessary for knowing virtue and being virtuous, and the outputs of sympathy are dependent on having the right character and disposition, then we have a way in which Hume’s view appears to be circular: there is a sense in which sympathy presupposes virtue. For the sympathizer must already possess the right disposition and character—be virtuous, in effect—in order to be able to know virtue and be virtuous.

Aspect (3) of the motivation objection is the claim that, because empathy is neutral about motive, using it to define or explain certain aspects of morality in a non-circular way is problematic. Although Hume is not attempting to define or explain aspects of morality in terms of sympathy, for example what virtue is or what being virtuous consists in, it is nonetheless possible to use the problem facing him to introduce the notion of circularity, for it shows how aspect (3) will apply in Chapters 3 and 4 to Smith and Slote, who do use sympathy (or empathy, in Slote’s case) to define or explain aspects of morality. For if the outputs of the sympathizing depend on the motives of
the sympathizer, then we can see how using these outcomes to define virtue will be circular: it will presuppose what it attempts to explain, namely virtue.

2.4.4 Summary

In this section I argued that Hume faces aspect (2) of the motivation objection in relation to moral judgement. I argued that the sympathizer’s motives—his disposition and character—mean that sympathizing cannot guarantee that he makes a genuinely moral judgement, for only those with a certain disposition and character are able to reliably discern moral sentiments from non-moral sentiments. A good judge, however, is much more easily able to distinguish between his moral sentiments and his situated sentiments, so he will be more likely to make genuine moral judgements. I argued that Hume could, however, accept this. But I then argued that he faces a more fundamental problem: why would a good judge ever elect to privilege the moral sentiments, felt from the common point of view, over his own, situated sentiments? For his moral sentiments to have any influence on him, that is, to motivate him to be moral, he must not only be able to discern the moral sentiments on a regular basis, but he must also be motivated to act in accordance with those sentiments. We saw that a good judge could potentially be motivated to privilege his moral sentiments and act according to them because he desires the pleasure caused by feeling approval. But we then saw that the good judge could feel more pleasure from approving from his own, situated perspective, in which case he would surely be more likely to act according to his situated sentiments as opposed to his moral ones. I considered the possibility of a general motive to be moral, and concluded that this option was not open to Hume, because he denies that such a motive exists independently of sympathy. Lastly, I introduced the notion of circularity that will feature in the next two chapters.
2.5 Conclusion

In this chapter I introduced the first of the three constitutive views of empathy’s role in morality that I consider in the thesis. According to Hume’s sentimentalist theory, sympathy constitutes morality insofar as it functions as a mechanism that is necessary for the production of compassion and is necessary for moral judgement. In section 2.2 of the chapter I introduced Hume’s notion of sympathy, and argued that his sympathy concept includes a limited form of HLE that involves the imagination. In section 2.3 I outlined and analysed sympathy’s role in the production of compassion and benevolence, and applied aspect (2) of the motivation objection. I argued that imaginative sympathy, as I called Hume’s limited form of HLE, although necessary for compassion, is not sufficient for it. It cannot guarantee that compassion is produced when we sympathize because what we feel as a result of sympathizing depends on the motivations—the disposition and character—of the sympathizer.

In section 2.4 I applied aspect (2) of the motivation objection to Hume’s theory of moral judgement. I argued that the motives of the sympathizer mean that sympathizing cannot guarantee that, if we sympathize, we will make a genuine moral judgement. Moreover, I argued that Hume’s theory cannot provide us with an adequate explanation of why we should privilege the moral sentiments we feel from the common point of view over our situated sentiments and act in accordance with these, rather than our situated sentiments. Finally, I introduced the notion of circularity that will appear in Chapters 3 and 4. Hume’s view entails that knowing virtue and being motivated to act for the good of others both depend on the outcomes of the sympathy mechanism, which in turn depend on the motivations of the sympathizer.
3 Adam Smith’s Sympathy

3.1 Introduction

In this chapter I turn to the second of the three constitutive views of empathy’s role in morality, the sympathy-based moral theory of Adam Smith. Smith was Hume’s contemporary and friend. His moral theory is sympathy-based, and sympathy plays a constitutive role, but Smith’s sympathy concept differs from Hume’s. For Smith, sympathy requires the perspective-taking forms of HLE I described in Chapter 1 (§1.2). Sympathizing is, for Smith, as it was for Hume, foundational for morality because it is necessary for moral evaluation and for compassion. Sympathy also serves to explain what being virtuous consists in: two distinct sets of virtues are founded, Smith claims, on the effort to sympathize.

In this chapter I argue that Smith faces all three aspects of the motivation objection. These are: (1) reliability: empathy processes are not reliable insofar as they do not always lead to the right results; (2) function: empathy is not sufficient for various functions; (3) circularity: using empathy’s outputs to define or explain certain aspects of morality in a non-circular way is problematic.

The chapter is set out as follows. In section 3.2 I introduce Smith’s concept of sympathy. I argue that sympathy is the outcome of a HLE process, similar to the one described in Chapter 1 as in-his-shoes-imagining, not the process itself. I argue that what Smith means by sympathy can only be produced by such a process, and not by a low-level process like contagion. I argue that there are at least two possible outcomes that sympathy (qua outcome) can give rise to: compassion and approbation. I also defend the view that Smith is a proto-simulationist, an early exponent of what is now referred to as
simulation-based emotion mindreading, against a new interpretation of Smith’s sympathy concept advanced by Bence Nanay.

Sections 3.3, 3.4 and 3.5 explicate three fundamental constitutive functions of sympathy and the process behind it, and apply the relevant aspects of the motivation objection to each. In section 3 I apply aspects (1) and (2) to Smith’s claim that the process behind sympathy is sufficient for compassion. Smith faces aspect (1) as follows. He claims that the HLE process is the only way we can know how another feels. He also claims that compassion arises as the result of another person’s misery and suffering. It is plausible to assume, therefore, that knowing that the other is suffering is necessary for compassion. However, I argue that HLE processes are unreliable, frequently giving rise to incorrect attributions of mental states to others. If compassion is dependent on making a correct attribution of suffering to the other person, then HLE will not be sufficient for compassion. I argue that Smith can accept this, however, given his examples of so-called ‘illusive’ sympathy, in which compassion can be felt in the absence of suffering. But aspect (2) of the motivation objection arises, even if an attribution of suffering is not necessary for compassion. Feeling compassion is dependent on HLE because it is only by putting ourselves in the other’s shoes that we can come to feel the ‘analogous’ passion that gives rise to our compassion. Compassion therefore depends on whether or not we would suffer in the other person’s situation. For if we would not, we will not feel sympathy, and will not thence feel compassion. Moreover, this form of HLE will not necessarily result in compassion even if the sympathizer would suffer in the other’s shoes. For there is evidence to show that our disposition and character can determine whether or not we feel personal distress or compassion as a result of this form of HLE. Smith could respond that his sympathy concept includes another form of HLE, imagining being the other person, which can
give rise to compassion and does not require that we must feel as the other feels if we were in his shoes. In addition, there is empirical evidence that this form of HLE is more likely to lead to compassion than in-his-shoes-imagining. Smith’s text does contain examples of compassion being caused by both forms of HLE. But I argue that the weight of evidence is in favour of compassion being caused by in-his-shoes-imagining, and that we ought to deem this to be Smith’s considered view, for it is this that forms the basis of his account of moral evaluation, the function of sympathy I turn to in section 3.4.

In section 3.4 I argue that Smith faces aspect (2) of the motivation objection in relation to moral judgement. Smith argues that moral judgement requires that the spectator put himself in the other’s shoes and imagine if he would be motivated to act as the actor does. If he would be, he will sympathize, and also approve, because sympathy entails approval. However, I argue that HLE is not sufficient for moral judgement because our dispositions and characters mean we can be motivated to make false predictions about how we would feel in the other’s situation. Our moral judgements of others can therefore be affected by our motives—we can sympathize and approve (or fail to sympathize and disapprove) of others based on what we falsely imagine we would feel in the other’s shoes. I argue that Smith could respond by invoking the device of the impartial spectator. This would ensure that our moral judgements are not affected by our self-evaluations and are thus impartial. Making moral judgements of others requires firstly adopting the perspective of an impartial spectator and making judgements of self-approval. If the impartial spectator approves of our conduct, then we can legitimately employ this as the standard against which to measure the conduct of others. But this requires spectators to have a reason (a motive) to adopt the perspective of the impartial spectator and judge themselves
according to his prescriptions. Smith could argue that we seek the sympathy and approval of the impartial spectator because it is pleasurable. However, agents could surely obtain pleasure from the sympathy and approval of those who do not represent the impartial spectator. So Smith must provide another reason why we are motivated to HLE with, and sympathize with, the impartial spectator. Smith rejects self-love and benevolence as motivations for adopting the perspective of the impartial spectator. I argue that, on Smith’s view, it seems as though we seek sympathy with the impartial spectator because we value being virtuous.

In section 3.5 I argue that Smith faces aspect (3) of the motivation objection in relation to his explanation of virtue in terms of sympathy: his explanation of virtue in terms of sympathy is circular. For Smith, virtue involves being fully sensitive to both the original and the sympathetic feelings of others, and having total command of our own feelings so that we are able to adjust them accordingly. The qualities of sensibility and self-command ensure that our sentiments and conduct will always be at the pitch the impartial spectator can most fully sympathize with, and therefore approve of. In order to establish whether or not his sentiments need adjusting, an agent must first of all adopt the perspective of the impartial spectator. However, I argue that doing this must require virtue, because the impartial spectator represents an objective moral standard that we must choose to adopt. I argue that Smith’s attempt to explain virtue as the disposition to sympathize with the impartial spectator is therefore circular, for sympathizing with the impartial spectator itself requires virtue. I consider a possible response Smith could make to this objection: putting oneself in the impartial spectator’s shoes is habitual, and as such does not require virtue. But I argue against this that, in order to become a habit, putting oneself in the impartial spectator’s shoes would have to be practiced, which in turn requires that the agent is aiming at virtue.
3.2 Sympathy

3.2.1 Introduction

In this section I examine the nature of Smith’s sympathy. I argue that sympathy is the outcome of an imaginative process, not the process itself. I describe the process, and argue that sympathy can only be produced by such a process, which involves imaginative perspective-taking, and not a low-level process like contagion. I argue that there are at least two possible outcomes that sympathy can itself give rise to. These are passions felt for others, and a sentiment of approbation. I also argue that Smith’s is an early simulation view of emotion mindreading, according to which we put ourselves in another’s shoes in order to know what they are feeling. I argue that one such construal of simulation can plausibly capture instances of sympathy, and defend the claim that Smith is a proto-simulationist from an objection advanced by Bence Nanay. I also argue that this new interpretation of Smith’s sympathy is insufficient.

For Smith, as well as Hume, ‘sympathy’ is a fundamental principle in human nature. Smith begins The Theory of Moral Sentiments (Smith, 1759/1976) by pitting himself firmly against Hobbes and Mandeville who considered man’s primary motive to be one of selfishness. He does this by stating his belief that man is, in actual fact, essentially compassionate in nature; we care about the well-being of others for its own sake, independently of our own interest, or any gain to ourselves. (TMS, p. 9, I.i.1.1) Indeed, claims Smith, even ‘the greatest ruffian, the most hardened violator of the laws of society, is not

---

63 Henceforth abbreviated to ‘TMS’. Citations include both page and section numbers.
64 Smith makes an extensive critique of the moral systems of Hobbes and Mandeville in Part VII of TMS.
altogether without’ compassion. On Smith’s view, compassion is ‘the emotion which we feel for the misery of others, when we either see it, or are made to conceive it in a very lively manner.’ (TMS, p. 9, I.i.1.1) Like Hume, Smith does not consider sympathy to be an emotion. It is, rather, the term Smith uses to denote our ‘fellow-feeling with any passion whatever’. (TMS, p. 10, I.i.1.5)

Pity and compassion are words appropriated to signify our fellow-feeling with the sorrow of others. Sympathy, though its meaning was, perhaps, originally the same, may now, however, without much impropriety, be made use of to denote our fellow-feeling with any passion whatever. (TMS, p. 10, I.i.1.5)

Sympathy (or fellow-feeling) can occur with emotions other than misery; we can also, according to Smith, feel joy with others when they feel joy, grief with them when they feel grief, happiness with them when they are happy, and so on. (TMS, p. 10, I.i.1.4)

Neither is it those circumstances only, which create pain or sorrow, that call forth our fellow-feeling. Whatever is the passion which arises from any object in the person principally concerned, an analogous emotion springs up, at the thought of his situation, in the breast of every attentive spectator. (TMS, p. 10, I.i.1.4)

3.2.2 Outcome vs. Process

A notable difference between Hume and Smith is that for the former, sympathy is a process, a mechanism of converting an idea into an impression, whereas for the latter, it is the outcome of an imaginative process. Hume did, of course, refer to outcomes of this conversion process using the term ‘sympathy’; extensive and limited sympathy are, as we saw,

---

65 Pity and compassion are, for Smith, as they were for Hume, synonymous: they are terms used interchangeably to describe what we feel as a result of sympathy or fellow-feeling with another’s misery.
66 Smith uses ‘sympathy’ and ‘fellow-feeling’ interchangeably. I will only use the term ‘sympathy’ in this chapter.
both outcomes of the sympathy mechanism. Calling sympathy an outcome is not the usual position taken by commentators on Smith: most consider his sympathy to be the imaginative process itself, that of imagining oneself in another’s situation. For example, Charles Griswold (2010, p. 65) argues that sympathy is a process of ‘conceiving how we should feel in the like situation.’ Stephen Darwall (1999, p. 142) has a reading that is closer to my own, however. For Darwall, Smith’s sympathy is ‘an imaginative sharing of the agent’s motive or the patient’s feeling from his point of view.’ (Ibid.)

It is clear from what Smith says that sympathy is the outcome of an imaginative process, rather than the process itself. It is by ‘changing places in fancy’ (‘fancy’ is a term Smith uses interchangeably with imagination) with someone else that we come to sympathize with them. (TMS, p. 10, I.i.1.3) The imaginative process results in our being ‘in sympathy’ with another person. It is important to recognise that sympathy is an outcome because, as we will see below in §3.4, a primary function of sympathy is to act as a standard against which we can evaluate the propriety of our own actions, opinions and sentiments, as well as those of others. The means by which we discover whether or not actions or sentiments are appropriate or inappropriate is by engaging in an imaginative act, the outcome of which tells us whether we are in sympathy with someone (or not). Discovering whether we would or would not sympathize gives us a standard against which to measure the propriety or impropriety of their actions, opinions and sentiments. The process of imagining and the result of the imagining—being ‘in sympathy’ with, or sympathizing with, another person—are distinct and separate.

Further evidence that sympathy is an outcome can be taken from Smith’s reference in various places in TMS to the Stoic concept of natural harmony. According to Raphael and Macfie (Smith, 1759/1976, introduction, p. 7) in their introduction to the Glasgow Edition of TMS, Smith was heavily influenced by Stoic thinking. The concept of a harmonious system advocated by the Stoics is the idea that there is a natural harmony and order to the universe. According to Smith, the moral sphere is a microcosm of the universe, and so operates according to an ordered and harmonious system, as the universe does. Or, at least, it does when it is working properly. We derive pleasure from observing this harmony when it exists. (Cf. Harrison, 1995, p. 101) When our feelings are in accord with those of others—when they are in harmony with them—we derive great pleasure and satisfaction from it, presumably because this is how things ought to be, according to the workings of the perfect and divine system. When there is discord, it is painful for us to observe it. Things are supposed to work in certain ways, and when they do not, we feel as though this is, in some way, ‘unnatural’. The second section of TMS states that ‘whatever may be the cause of sympathy, or however it may be excited, nothing pleases us more than to observe in other men a fellow-feeling with all the emotions of our own breast.’ (TMS, p. 13, I.i.2.2) And Smith claims later on in TMS that it is the pursuit of the pleasure we feel upon perceiving this harmony of feeling—in other words, upon perceiving that we sympathize with others—which motivates us to bring our own sentiments into line with those of others. (TMS, p. 22, I.i.4.7-8) Moreover, Raphael and MacFie (Smith, 1759/1976, introduction, p. 7) note that ‘the Stoics themselves applied the notion [of natural harmony] to society no less than to the physical universe, and used the Greek word sympatheia (in the sense of organic connection) of both.’ In light of this, it is reasonable to suggest that sympathy is an outcome, a state

of our emotions and feelings (and opinions and thoughts) being in harmony with those of others.

### 3.2.3 Sympathy and Imagination

Sympathy can only arise, Smith claims, as a result of an imaginative process. Although it may appear to arise via some other sort of process that does not involve the imagination, referred to in Chapter 1 (§1.2) as LLE, this is not the case. Some emotions—Smith singles out grief and joy—‘may seem to be transfused from one man to another, instantaneously,’ via what we might want to call some sort of LLE process. (TMS, p. 11, I.i.1.6, my italics)

Upon some occasions sympathy may seem to arise merely from the view of a certain emotion in another person. The passions, upon some occasions, may seem to be transfused from one man to another, instantaneously, and antecedent to any knowledge of what excited them in the person principally concerned. Grief and joy, for example, strongly expressed in the look and gestures of any one, at once affect the spectator with some degree of a like painful or agreeable emotion. A smiling face is, to every body that sees it, a cheerful object; as a sorrowful countenance, on the other hand, is a melancholy one. (TMS, p. 11, I.i.1.6)

Some commentators take Smith to be allowing for the possibility of LLE leading to sympathy in this paragraph. Griswold (2010, p. 65), for example, claims that ‘in the sixth paragraph [Smith] acknowledges the possibility that on some occasions sympathy may spread emotions instantaneously’ and that he therefore allows for ‘sympathy as contagion’. But LLE or ‘sympathy as contagion’ is not sufficient for sympathy. As Smith goes on to make clear in subsequent paragraphs, we have to know what caused a passion in order to be able to sympathize with the person feeling it, or the person on the receiving end of it. Simply seeing the outward signs of a passion is not enough to give rise to sympathy. (TMS, pp. 11, I.i.1.6; 12, I.i.1.10) As Smith
explicitly claims, sympathy ‘does not arise so much from the view of the passion, as from that of the situation which excites it.’ (TMS, p. 12, I.i.1.10)69

Smith argues that, although the very appearance of joy (‘a smiling face’) or grief (‘a sorrowful countenance’) is enough to ‘inspire us with some degree of the like emotions’ (TMS, p. 11, I.i.1.9), this is not because the other’s passion is ‘transfused’ into us. It is, Smith claims, because merely seeing the effects of grief and joy in another person is sufficient to give us ‘the general idea of some good or bad fortune that has befallen’ that person. (Ibid.) In other words, seeing a person’s smiling or sad face is enough to give us some idea of what might have caused these particular passions, namely good or bad fortune. But an idea of the cause of the other’s passion is still therefore needed in order for us to sympathize with the person experiencing it. In the case of grief and joy, seeing the effects—a smiling face or a sorrowful countenance—is sufficient to give us some general idea of the cause, which can give rise to some degree of sympathy. But we can only sympathize ‘perfectly’, even with grief and joy, when we are fully informed as to what the exact circumstances were that caused the person to feel those passions. (TMS, p. 11, I.i.1.9) That we must know the cause of a passion to sympathize fully with it is confirmed by the following passage.

Even our sympathy with the grief or joy of another, before we are informed of the cause of either, is always extremely imperfect. General lamentations, which express nothing but the anguish of the sufferer, create rather a curiosity to inquire into his situation, along with some disposition to sympathize with him, than any actual sympathy that is very sensible. The first question we ask is, What has befallen you? Till this be answered, though we are uneasy both from the vague idea of his misfortune, and still more from torturing ourselves with conjectures about what it may be, yet our fellow-feeling is not very considerable. (TMS, p. 11, I.i.1.9)

---

69 This is in marked contrast to Hume who, as we saw in Chapter 2, argued that sympathy can arise from observing the outward effects of a passion.
Further evidence that knowledge of the cause is necessary for sympathy can be provided by considering one of Smith’s examples, involving the passion of anger. He argues that, if A is angry with R, and S does not know why, S cannot sympathize with A. He will, however, be able to sympathize with R. Nature, Smith claims, ‘teaches us to be more averse to enter into’ anger until we are aware of its cause. In fact, until we are made aware of the cause, we are ‘disposed rather to take part against’ the angry person, A. (TMS, p. 11, I.i.1.8) If we are unaware of the cause of A’s anger, we are more likely to be able to ‘enter into’, and therefore sympathize with, the fear and resentment of the person A is angry with, viz. R.

Imagine that A is a father whose son has been assaulted by R. If S sees A’s angry behaviour towards R, and does not know why A is angry, Smith claims that he will be able to sympathize with R’s fear because he will know why R is feeling fear—he is being confronted by a very angry man—and will thus be able to imagine being frightened for that same reason, because he knows the cause of R’s fear: A’s anger. Conversely, S will be limited to imagining that he is angry, for no particular reason, if all he can see are the outward signs of A’s anger.

So far in this section I have defended the claim that Smith believes that sympathy can only arise as a result of an imaginative process. We have seen that knowledge of the cause is required for sympathy. If knowledge of the cause is necessary for sympathy, as Smith claims (e.g. TMS, p. 11, I.i.1.9) then it seems unlikely that sympathy would arise out of some sort of LLE process, because knowledge of the cause is not going to be necessary for S to feel something similar to what A feels as a result of LLE. However, it could still

---

70 A is the ‘actor’, R is the ‘recipient’, and S is the ‘spectator’.
be maintained that, given what we know now after years of research into empathy phenomena, LLE could actually result in S’s being in sympathy with A’s emotion, particularly if this is a ‘basic’ emotion state like sorrow.71 But LLE cannot be sufficient for Smithian sympathy. For, as I explain below (§3.4) it has to be possible for us to fail to sympathize with another person’s passion. This is an important part of Smith’s story, as experiencing a lack of sympathy with another person’s sentiment is what enables us to disapprove of him. If sympathy were the result of LLE, which is triggered automatically, then it could not be the case that S ever fails to sympathize, for when faced with A feeling some emotion, S would automatically sympathize with it, coming to feel what he feels.

3.2.3.1 Physical Symptoms

It is easy to see why some people interpret Smith’s examples of sympathy as instances of LLE. Goldman (2006, p. 17), for example, has suggested that Smith was an early detector of motor mimicry, which on his view is a form of LLE.72 He quotes the following example of what he claims is one of Smith’s observations of motor mimicry.

When we see a stroke aimed and just ready to fall upon the leg or arm of another person, we naturally shrink and draw back our own leg or our own arm; and when it does fall, we feel it in some measure, and are hurt by it as well as the sufferer.

(TMS, p. 10, I.i.1.3)

The paragraph from which this quotation is taken is replete with what Smith takes to be ‘many obvious observations’ like this one. (TMS, p. 10, I.i.1.3) And this example does, indeed, look like what we would today consider to be an instance of motor mimicry. However, Smith insists that these are the results of HLE. The above quotation begins as follows.

71 See e.g. Goldman (2006, p.147). Also see Ekman (1999) for a list of so-called ‘basic’ emotions.
72 See Oberman & Ramachandran (2007) for a similar view.
That this is the source of our fellow-feeling for the misery of others, that it is by changing places in fancy with the sufferer, that we come either to conceive or to be affected by what he feels, may be demonstrated by many obvious observations, if it should not be thought sufficiently evident of itself. When we see a stroke aimed [...] (TMS, p. 10, I.i.l.3)

In this paragraph Smith is describing the physical effects that can arise as a result of changing places in imagination with the sufferer. There are several other instances in the same paragraph, all exemplifying how we can experience physiological effects as a result of putting ourselves in another's shoes. When watching an acrobat on a slack rope, a crowd will ‘naturally writhe and twist and balance their own bodies, as they see him do, and as they feel that they themselves must do if in his situation.’ (TMS, p. 10, I.i.l.3) When we look upon ‘the sores and ulcers which are exposed by beggars in the streets,’ we feel ‘an itching or uneasy sensation’ in the corresponding part of our own bodies. (Ibid.) Bence Nanay (2010, p. 92) claims that this is an instance of emotional contagion. However, Smith is clear that this outcome arises via changing places in our imaginations with the beggars, not by contagion, as the following quotation shows.

The horror which they conceive at the misery of those wretches affects that particular part in themselves more than any other; because the horror arises from conceiving what they themselves would suffer, if they really were the wretches whom they are looking upon, and if that particular part in themselves was actually affected in the same miserable manner. (TMS, p. 10; I.i.l.3)

The ‘very force of the conception’ is enough to make us have physical symptoms in our own bodies in this case. (TMS, p. 10; I.i.l.3) Imagining how we would feel in their shoes, afflicted as they are, can cause us to have the beggars’ physical symptoms. This would constitute our sympathizing with them. Looking upon a person with sore eyes can cause us to have sore eyes
ourselves ‘from the same reason’, that is, from putting ourselves in their shoes. (Ibid.) It is clear that, contra commentators like Nanay, these so-called instances of LLE are, in fact, caused by HLE.

3.2.3.2 Bodily and Imaginative Passions

Further evidence that sympathy is produced only by the imagination can be seen in the section in TMS where Smith makes a distinction between bodily passions and passions arising from the imagination. We find it very difficult to sympathize with bodily passions, Smith claims. Take hunger, for example. Because we cannot make ourselves feel hungry to the extent that the other person is hungry simply by putting ourselves in another’s shoes, we cannot sympathize to any great extent. (TMS, p. 27, I.ii.1.1) We can sympathize to a very limited extent, however. When A has a good appetite at dinner, S can imagine being similarly hungry at mealtimes, because it is natural for all healthy people to be hungry at particular times of day. (TMS, p. 27, I.ii.1.1) But S cannot literally grow hungry to the extent that A is hungry merely by imagining being hungry, according to Smith, so he cannot sympathize to any great extent with A’s hunger.

Similarly, S cannot sympathize to any great extent with A if he is in severe pain, also a bodily passion. Smith had earlier claimed that we can sympathize with another’s pain. But he recites one of his own examples (seeing the stroke aimed at another’s leg) and reminds the reader that, although we can feel hurt by it as well as the sufferer, our ‘hurt, however, is, no doubt, excessively slight.’ (TMS, p. 29, I.ii.1.5) Furthermore, Smith claims, this is the case with ‘all the passions which take their origin from the body: they excite either no sympathy at all, or such a degree of it, as is altogether disproportioned to the violence of what is felt by the sufferer.’ (Ibid.) So our own sympathy with the beggars’ physical afflictions will be slight, because it
is only possible that we can feel a faint replica of their physical symptoms. Nevertheless, we can still sympathize with them, albeit to a limited extent.

Conversely, passions that arise from the imagination can be sympathized with much more fully. We can feel these as the other feels them, Smith claims. Passions of the imagination include grief, fear and ‘disappointment in love or ambition’. (TMS, pp. 28, I.ii.1.2; 29, I.ii.1.6) Spectators can easily imagine feeling grief, fear and disappointment. This means they can sympathize much more easily with others when they are feeling these passions. As Smith puts it, ‘the frame of my body can be but little affected by the alterations which are brought about upon that of my companion: but my imagination is more ductile, and more readily assumes, if I may say so, the shape and configuration of the imaginations’ of others. (TMS, p. 29, I.ii.1.6)

3.2.4 Sympathy and Simulation

In this section I argue that Smith’s view is proto-simulationist. That is, the process underlying Smith’s sympathy is an early form of simulation. It is common to read Smith in this way. According to Gordon (1995, p. 741), for example, Smith’s is an early version of how we ‘psyche out’ another’s emotions by putting ourselves in their shoes and then projecting how we would feel onto the other person. That this does appear to be Smith’s view seems clear from the second paragraph of TMS, quoted below.

As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation. Though our brother is on the rack, as long as we ourselves are at our ease, our senses will never inform us of what he suffers. They never did, and never can, carry us beyond our own person, and it is by the imagination only

73 Also see TMS (p. 21, i.i.4.6).
that we can form any conception of what are his sensations. Neither can that faculty help us to this any other way, than by representing to us what would be our own, if we were in his case. It is the impressions of our own senses only, not those of his, which our imaginations copy. By the imagination we place ourselves in his situation, we conceive ourselves enduring all the same torments, we enter as it were into his body, and become in some measure the same person with him, and thence form some idea of his sensations, and even feel something which, though weaker in degree, is not altogether unlike them. His agonies, when they are thus brought home to ourselves, when we have thus adopted and made them our own, begin at last to affect us, and we then tremble and shudder at the thought of what he feels. For as to be in pain or distress of any kind excites the most excessive sorrow, so to conceive or to imagine that we are in it, excites some degree of the like emotion, in proportion to the vivacity or dullness of the conception. (TMS, p. 9, I.i.1.2)

According to the simulation theory, as we saw in Chapter 1 (§1.3.2), when we attribute mental states to others, we do so by simulating them. That is, we put ourselves in their mental shoes and run a simulation, the output of which is some pretend mental state that we then attribute to that person. A simulationist interpretation of this paragraph would be as follows. Our brother is on the rack, and the way we come to know how he feels is by ‘placing ourselves in his situation’, and imagining ‘ourselves enduring all the same torments’. We thereby ‘form some idea of his sensations and even feel something which, though weaker in degree, is not altogether unlike them.’ (TMS, p. 9, I.i.1.2) The final stage would be our attribution of these states to our brother. *Figure 8* below is my interpretation of Smithean sympathy. It shows two possible consequences of sympathy: emotions that we feel for others, and the sentiment of approbation. The sympathizing occurs at Stage 2.
3.2.4.1 Sympathy and E-imagination

Goldman’s notion of E-imagination provides a plausible interpretation of the process Smith is describing in this paragraph. In Chapter 1 (§1.3.2) we saw that, on Goldman’s view, E-imagining subserves high-level simulation-based mindreading. Goldman (2006, pp. 47-48) draws a distinction between what he calls, respectively, suppositional imagining (S-imagining) and enactment
imagining (E-imagining). E-imagining involves enacting the mental state of another person in imagination. Enacting being in their mental state is recreating or replicating that state in ourselves. *(Ibid., p. 283)* We ‘conjure up a state that feels, phenomenologically, rather like a trace or tincture’ of, for example, an emotion or feeling like elation. *(Ibid., p. 47)* S-imagining is different. When a person S-imagines, they ‘suppose, assume, or hypothesize’ that some state of affairs is the case. In other words, in the case of elation, they assume the truth of the proposition ‘I am elated’. *(Ibid.)* The outcome or product of S-imagining is a supposition or assumption that a state of affairs obtains. The outcome or product of an act of E-imagining, on the other hand, can be any number of mental states, including emotion states, for it is possible to enact many mental states in our imaginations. *(Ibid.)*

In *Figure 8*, sympathy is the output of the imaginative process. It is a felt version of what we imagine we would feel in the other person’s situation that we then attribute to the other person. We thereby come to know how they feel. A simulation process like E-imagining can account for the outcomes that Smith claims are instances of sympathy. The advantage of being able to E-imagine anger, for example, according to Goldman, is that it can lead to S’s feeling an anger-like state (what Goldman calls *quasi*-anger), which is itself more likely to lead to further emotional responses than S-imagining. *(Goldman, 2006, p. 283)* Evidence for this can be taken from the literature on simulation and emotional responses to fiction. According to Gregory Currie and Ian Ravenscroft (2002, p. 20), for example, affective responses to fiction can be explained by positing the existence of imagined states that are desire-like or belief-like, but are not actual desires or beliefs. They are, however, sufficiently like these states to give rise to the kinds of
emotional responses that would result if the states were ‘real’ and not imagined.

Goldman (2006, p. 284) takes this further to include emotion-like states, for example quasi-anger. According to what Goldman calls ‘the E-imagination hypothesis’, ‘affective responses to fiction occur because fiction serves as a series of textual or theatrical props that fuel a viewer’s or a reader’s E-imagination into producing all sorts of surrogate states.’ (Ibid.) These surrogate states resemble their counterpart states sufficiently to produce the same affective responses that would arise if the states were real, not merely imagined. (Ibid.) So if I E-imagine feeling E, I will feel quasi-E, and can have the same emotional responses to the quasi-E as I would have to E. (Ibid.) If I merely S-imagine feeling E, I cannot experience quasi-E, and am less likely to feel any related emotions.

In the case of A, the angry man, the thought is that if S knows the cause of A’s anger, he is more easily able to E-imagine being angry because he can enact being in that state in his imagination for the same reason as A. If S can E-imagine being angry for the same reason that A is angry, he can feel quasi-anger. Moreover, he might then have a further emotional response, the same response he would have if he were feeling real anger. If S does not know the cause, and he is limited to S-imagining, then he cannot produce these pretend states and cannot sympathize. He is also less likely to feel any subsequent emotions.

3.2.4.1.1 Physical Symptoms

Earlier (§3.2.3.1) I argued that, according to Smith, physical symptoms arise as a result of in-his-shoes-imagining, and not some LLE processes. E-imagining might account for how this occurs as follows. The spectator
would put himself in the situation of the beggars, imagining how he himself would feel if he were in their situation, including having the same physical afflictions. He would E-imagine misery, feel quasi-misery, and then attribute this to the beggars. He might also feel quasi-feelings that would constitute the physical symptoms similar to those being felt by the beggar. Imagining having those sores might give rise to his feeling the same physical symptom. He could also then feel the horror that onlookers feel when they are confronted by beggars; this, Smith claims, would arise ‘from conceiving what they themselves would suffer, if they really were the wretches whom they are looking upon.’ (TMS, p. 10, I.i.1.3) The horror constitutes a passion we can feel for others on account of sympathizing with them.

3.2.4.1.2 Compassion

This interpretation can also account for our feeling compassion. When we put ourselves in the other person’s shoes, E-imagine sorrow or misery, feel quasi-misery, and attribute that misery to others, we will then feel compassion for the other person because we believe them to be suffering or miserable. As we saw at the beginning of §3.2, compassion is ‘the emotion which we feel for the misery of others, when we either see it, or are made to conceive it in a very lively manner.’ (TMS, p. 9, I.i.1.1) Compassion is an ‘original passion of human nature’, and as a result is capable of being felt by everyone, not just ‘the virtuous and humane’. Indeed, even the ‘greatest ruffian’ feels compassion when he sees it or is made to conceive it. (TMS, p. 9, I.i.1.1)

3.2.4.1.3 Approbation

The sentiment of approbation that is required for moral approval arises, on Smith’s view, as a result of HLE. As I will explain more fully in §3.4, Smith holds that our moral judgements involve feeling a sentiment of approbation. By putting ourselves in the other’s shoes, we not only come to know how the
other person feels, but also how we would feel in their situation. If the imagining results in our coming to feel as they do, we will achieve a correspondence of sentiments with the other person. If we sympathize fully with their sentiments, we will then feel a sentiment of approbation. The sentiment of approbation can only be produced if we achieve ‘complete sympathy’: a ‘perfect harmony and correspondence of sentiments’ is required for approbation. (TMS, p. 44, I.iii.i.3)

3.2.4.1.4 Bodily Passions

Earlier (in §3.2.3.2) we saw that bodily passions like hunger can be sympathized with insofar as we feel a weak version of what the other is feeling, but they can never be sympathized with to any great extent because it is impossible for us to feel the passion to the same degree as the sufferer. In other words, it is impossible for us to sympathize with them to any great extent—we cannot make ourselves as hungry as the sufferer. Passions of the imagination, on the other hand, can be more fully sympathized with because we can more easily imagine ourselves feeling them. From this it possible to see why, for Smith, sympathy is necessary for approval, and how the current interpretation of Smith’s view can account for this. On Smith’s view, ‘it is indecent to express any strong degree of those passions which arise from a certain situation or disposition of the body [...] because the company, not being in the same disposition, cannot be expected to sympathize’ with any strong degree of a bodily passion. (TMS, p. 27, I.ii.1.1) This means they (the company) will never approve of behaviour that expresses a strong degree of these passions. (TMS, p. 27, I.ii.1.1) As Smith puts it: ‘violent hunger, for example, though upon many occasions not only natural, but unavoidable, is always indecent, and to eat voraciously is universally regarded as a piece of ill manners.’ (TMS, p. 27, I.ii.1.1) There is, however, ‘some degree of sympathy, even with hunger’. (Ibid.) The situation is similar in the case of pain:
It is for the same reason that to cry out with bodily pain, how intolerable soever, appears always unmanly and unbecoming. There is, however, a good deal of sympathy even with bodily pain. (TMS, p. 29, I.ii.1.5)

The current account of sympathy captures the fact that there can be some degree of sympathy with hunger and pain, but no approval of actions manifesting these bodily passions. For we can E-imagine being hungry and in pain, and it is plausible that we can feel a degree of quasi-hunger or quasi-pain as a result, which would suffice for making an attribution of those states to the other person. Moreover, it explains why we will not approve of violent expressions of these passions, for we cannot sympathize entirely with them. Because complete sympathy is necessary for approval, and we cannot E-imagine ourselves into feeling complete sympathy for bodily passions, we cannot approve of displays of such passions.

3.2.4.2 Nanay’s Objection

A challenge to the simulationist reading of Smith’s sympathy has been made recently by Bence Nanay (2010). Nanay (ibid, p. 88) argues that Smith’s sympathy cannot be simulation because it ‘does not entail any correspondence between the mental states of the sympathizer and the person she is sympathizing with.’ Those who consider Smith’s sympathy to be simulation, he argues, wrongly assume that it does imply such a correspondence. For Smithian sympathy to be an instance of simulation, Nanay argues, it ‘must be a correct guide to the other person’s mental states.’ (Nanay, 2010, p. 94) To be a correct guide to the other person’s mental states, the simulation process must always result in S making correct attributions. In order to make correct attributions, there must always be a match between what the target feels, and what S comes to feel by simulating the target.

75 On Nanay’s (2010) view, ‘sympathy’ describes the imaginative process, not the outcome of this process.
Smith has examples of sympathy in which no correspondence occurs. In such cases we:

feel for another, a passion of which he himself seems to be altogether incapable; because, when we put ourselves in his case, that passion arises in our breast from the imagination, though it does not in his from the reality. (TMS, p. 12, I.i.1.10)

Nanay argues that the existence of such instances, in which there is no correspondence of sentiments between S and the target, shows that sympathy cannot be simulation. For if we can still sympathize in cases in which there is no correspondence, and simulation requires correspondence, then sympathy cannot be simulation. Two examples of sympathy without correspondence are sympathy with the dead and sympathy with the insane. The relevant sections of TMS are quoted below. The first quotation describes sympathy with the insane, and the second sympathy with the dead.

Of all the calamities to which the condition of mortality exposes mankind, the loss of reason appears, [...] by far the most dreadful [...] But the poor wretch, who is in it, laughs and sings perhaps, and is altogether insensible of his own misery. The anguish which humanity feels, therefore, at the sight of such an object, cannot be the reflection of any sentiment of the sufferer. The compassion of the spectator must arise altogether from the consideration of what he himself would feel if he was reduced to the same unhappy situation, and, what perhaps is impossible, was at the same time able to regard it with his present reason and judgement. (TMS, p. 12, I.i.1.12)

We sympathize even with the dead [...] The idea of that dreary and endless melancholy, which the fancy naturally ascribes to their condition, arises altogether from our joining to the change which has been produced upon them, our own consciousness of that change, from our putting ourselves in their situation, and from our lodging, if I may be allowed to say so, our own living souls in their inanimated bodies, and thence conceiving what would be our emotions in this case. (TMS, p. 13, I.i.1.13)
Smith describes sympathy with the dead as occurring as the result of an ‘illusion of the imagination’. (TMS, p. 13, I.i.1.13) The imagining is illusory because it involves S imagining how he would feel in a situation that he knows he cannot, in reality, find himself in. In the case of sympathy with the man who has lost his reason, the illusion of the imagination involves S imagining how he would feel if he himself had lost his reason but was simultaneously able to contemplate his situation ‘with his present reason and judgement.’ (TMS, p. 12, I.i.1.12)

3.2.4.3 Response 1

Nanay’s argument rests on his assumption that, to count as an instance of simulation, correct attributions must always be made. That is, simulations must always produce correct attributions. This is not the case, however. Simulation theorists like Goldman (2006, p. 150), for example, hold that a process can count as a simulation even though it does not result in a correct attribution, because it is not the correctness or otherwise of the attribution that matters. It is, in actual fact, ‘the ostensible purpose or function’ of the process that makes it an instance of simulation, and not the reliability or correctness of the attributions that are then made. So we could still say that Smith’s is a proto-simulationist view, if its intended function is to facilitate third-person emotion attributions, regardless of whether it actually does so in every case. This means that sympathy with the dead and insane could still be simulation even though there is no correspondence of sentiments.

3.2.4.4 Response 2

Nanay’s view implies that the sympathizer’s state must be an exact replica of the target’s state if it is to count as a correct attribution. However, on Smith’s view, the feeling or emotion state that a spectator feels need only be similar to that of the target to count as sympathy; it does not have to be identical or exactly corresponding. Sympathizing requires putting ourselves in another’s
shoes and coming to feel something which, ‘though weaker in degree’ is ‘not altogether unlike’ what the other is feeling. (*TMS*, p. 9, I.i.1.2) We attribute this to the other person, thereby coming to know how he feels. Smith appears to have a loose definition of the term ‘analogous’. When he uses the term to describe the emotion which arises ‘at the thought of another person’s situation’ (*TMS*, p. 10, I.i.1.4), it does imply that the spectator feels an identical emotion to the one the actor is experiencing. But Smith also claims that ‘analogous’, when applied to a passion, can mean simply that one passion bears similar phenomenological characteristics to another. (*TMS*, p. 324, VII.iii.3.13) On Smith’s view, emotions can share ‘general features’ which mark them out as emotions of certain types, and this is sufficient for them to be considered ‘analogous.’ Anger, for example, has certain distinguishing features and can appear in different guises. But it is nevertheless always discernible as a ‘species’ of anger. (*TMS*, p. 324, VII.iii.3.13) And Smith uses ‘similar’ as well as ‘analogous’ to describe outcomes of in-his-shoes-imagining. He claims that this form of HLE leads to feeling ‘emotions similar to what [the other] feels.’ (*TMS*, p. 22, I.i.4.8) Given this non-rigid definition of ‘analogous’, it seems plausible to suggest that, for Smith, the spectator need only feel, as a result of putting himself in the target’s shoes, something reasonably similar to what the target feels. He can then attribute this to the target, and thence feel any subsequent emotions. This leaves room for the output of the imagining to be ‘corresponding’, even though it is not exactly analogous, which means an attribution could still be correct.

3.2.4.5 Response 3

Nanay does not provide an alternative that is preferable to simulation as an interpretation of Smith’s sympathy. Although this is insufficient by itself to prove that Smith’s sympathy is simulation, or to show definitively that Nanay’s is the incorrect interpretation, it does provide evidence in favour of
the simulation alternative over Nanay’s construal. Nanay (2010, p. 97) argues that sympathy is best regarded as a ‘very basic, quasi-automatic process that happens to us without any concern about whether we are “getting it right” what mental state the other person is in’. It is ‘a simple imaginative process: imagining ourselves in a certain situation—the situation we take the other person to be in, which may or may not be the situation she takes herself to be in.’ (Ibid.) Nanay cites Smith’s examples of physical symptoms as support for his claim that sympathy must be such a quasi-automatic process. In fact, he claims, if the instances of physical sympathy that we saw earlier are to count as instances of sympathy, then ‘sympathy can only be a visceral, quasi-automatic reaction of imagining ourselves in a certain situation—the situation we perceive or believe someone else to be in’. (Ibid., my italics) But a simulation alternative could plausibly account for such physical symptoms, as we saw earlier. The spectator could feel a similar physical symptom to the target as a result of E-imaging.

Nanay is right that the imaginative process can be spontaneous and non-deliberate: Smith says explicitly at one point that the momentary change of places necessary for sympathy can happen ‘instantaneously’. (TMS, p. 22-23, I.i.4.9-10) But he is wrong to claim that the imaginative process is only quasi-automatic. It can also be deliberate. Smith claims that, at least on some occasions, we have to deliberately take pains to recreate another’s circumstances in order to sympathize. The spectator must ‘endeavour, as much as he can, to put himself in the situation of the other [...] and strive to render as perfect as possible, that imaginary change of situation upon which his sympathy is founded.’ (TMS, p. 21, I.i.4.6) Given that the imaginative process can be deliberate and spontaneous, Nanay’s new interpretation appears too narrow and does not encompass what Smith had in mind.
3.2.5 Summary

In this section I introduced Smith’s concept of sympathy. I argued that sympathy is the outcome of an imaginative process, not the process itself. I argued that sympathy is only produced by the imagination, and is not the result of some form of LLE process. I argued that there are at least two possible outcomes that sympathy can itself give rise to: passions felt for others and a sentiment of approbation. I argued that Smith’s is an early version of simulation-based emotion mindreading, and that one such simulation process that could plausibly underpin this is Goldman’s notion of E-imagination. I defended the claim that Smith is a proto-simulationist against an objection put forward by Bence Nanay, and argued that his alternative interpretation of Smith’s sympathy is unsatisfactory. In the next section I argue that Smith faces aspects (1) and (2) of the motivation objection in relation to his claim that the HLE process behind sympathy is sufficient for compassion.

3.3 Compassion

In this section I argue that Smith faces aspects (1) and (2) of the motivation objection in relation to his claim that HLE is sufficient for compassion. Aspect (1) is the reliability aspect: our motives mean that empathy processes are not reliable as they do not always lead to the right results. Aspect (2) is the function aspect: our motives mean that empathy processes are not sufficient for one of their purported functions, in this case causing compassion. I apply aspect (1) in section 3.3.1, and aspect (2) in section 3.3.2.

3.3.1 Aspect (1): Reliability

As we saw at the beginning of §3.2, compassion is ‘the emotion which we feel for the misery of others, when we either see it, or are made to conceive it in a very lively manner.’ (TMS, p. 9, I.i.1.1) It is an ‘original passion of human nature’ that is capable of being felt by everyone, not just ‘the virtuous and
humane’; even the ‘greatest ruffian’ feels compassion for others’ misery and suffering when he sees it or is made to conceive it. (TMS, pp. 9, I.i.1.1; 43, I.iii.1.1) We also saw that, on Smith’s view, the HLE process is the only way we can come to know how the other person feels. (TMS, p. 9, I.i.1.2)

As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation. (TMS, p. 9, I.i.1.2)

Given that compassion arises as the result of another person’s misery and suffering, it is reasonable to assume that, in order to feel compassion, the spectator must know that the other is suffering. This means he must make an attribution of suffering to the other by engaging in HLE. However, as I will now argue, the HLE process is unreliable, and frequently gives rise to incorrect attributions of mental states to others. If feeling compassion depends on making correct attributions of suffering to others, then HLE will not guarantee that we feel compassion because it cannot reliably produce correct attributions.

According to Goldman (2006, p. 40), when making emotion attributions to others, a simulator takes ‘one of her own first-order (pretend) states and imputes it (as a genuine state) to the target.’ In the case of emotion mindreading, a simulator would take the pretend emotion that he assumes the other person to have, E-imagine feeling it, and then impute it as a genuine emotion to a target. However, simulators must be sure to ‘quarantine’ their own idiosyncratic mental states from a simulation in order to make accurate attributions. (Ibid., p. 29) Quarantining in the case of emotion mindreading would require ensuring that our own emotions do not ‘seep into’ our simulation routine and ‘contaminate’ it. (Ibid.) If the pretend emotion that we E-imagine feeling is not shared by the target, then failing to
exclude it from our simulation will result in an inaccurate attribution of that emotion to the target.

Quarantine failure, also known in psychology as projection or egocentrism, can occur in at least three areas: knowledge, valuations and feelings. (Goldman, 2006, p. 165) It is well documented in the literature on simulation that a person’s knowledge, values and feelings can affect her attributions. The so-called ‘curse of knowledge’ is where the simulator fails to account for the fact that she possesses certain knowledge that the target does not. The so-called ‘endowment effect’ is another factor that can influence attributions. If the simulator is rich, for example, what he is ‘endowed with’, his riches, affect his ability to accurately predict the extent to which a poorer person will value something. (Goldman, 2006, p. 167) Finally, the occurrent feelings of the simulator can slip into his attribution. Goldman (2006, p. 167) cites as an example a study undertaken by Van Boven and Loewenstein (2003) in which volunteers were asked to predict how a group of hikers would feel if they got lost in some woods without food or water. Some volunteers made their predictions before doing strenuous exercise and others did so afterwards. The volunteers who made their predictions post-exercise, which had left them thirsty and hot, tended to predict that the hikers would feel more thirsty and hot than the volunteers who made their predictions before exercising. (Goldman, 2006, p. 168) If a correct attribution of suffering to the other person is required for compassion, and the process that gives rise to this attribution is unreliable because simulators are prone to quarantine failure, then we can conclude that HLE is not sufficient for compassion because it is unreliable.

---

76 See Goldman (2006, pp. 164-170) for further examples of projection.
3.3.2 Aspect (2): Compassion

Smith could respond by denying that any attribution of suffering is necessary for compassion. As we saw earlier (§3.2.4.2), his examples of illusive sympathy are cases in which compassion is felt for another person in the absence of a correspondence of sentiments. In these cases, the spectator feels compassion on the basis of what he imagines he himself would feel if he were in the sufferer’s shoes, and not on the basis of what the sufferer is actually feeling. The spectator feels ‘for another a passion of which he himself seems to be altogether incapable; because, when we put ourselves in his case, that passion arises in our breast from the imagination, though it does not in his from the reality’. (TMS, p. 12, I.i.1.10) In the two cases of illusive sympathy quoted earlier, the dead and insane are not suffering or miserable. The dead are, of course, insensate because they are dead; and the man who has lost his reason ‘laughs and sings’ and is ‘altogether insensible of his own misery.’ (TMS, p. 12, I.i.1.11) If no attribution is necessary, then it does not seem to matter whether the process underlying attributions is reliable.

However, aspect (2) of the motivation objection arises even if an attribution of suffering to the target is not necessary for feeling compassion for them. Feeling compassion is dependent on HLE because it is only by putting ourselves in the other’s shoes that we can come to feel the ‘analogous’ passion that gives rise to our compassion. To feel the analogous passion—the sympathy—the spectator must put himself in the target’s shoes and imagine how he himself would feel if he were in that person’s situation. Given that compassion is felt as a result of sympathy with sorrow, on Smith’s view, the ‘analogous’ passion the spectator must feel is grief or misery or sorrow. The sympathizer must therefore conclude as a result of putting himself in the other’s shoes that he would be miserable in that person’s
situation. Compassion thus depends on the spectator feeling misery or suffering in the other’s situation. If the spectator would feel something else—or he would not suffer or be miserable in the same situation—then he will not feel sympathy, and will therefore fail to feel compassion.

In-his-shoes-imagining looks insufficient for compassion because the disposition and character of the spectator will determine whether or not he would or would not feel misery in the other’s situation: it is reasonable to assume that not everyone will feel miserable or suffer in every situation. The cases of illusive sympathy imply that, on Smith’s view, our compassion is felt as a result of how we ourselves would feel, and seems altogether unrelated to what the other person actually feels. Making it the case that we can only feel compassion for others if we would feel miserable in their shoes seems like a very restrictive notion of compassion; it is surely the case that, as Lawrence Blum (1994, p. 176) notes, compassion arises from our consideration of the other’s distress, not our own.\textsuperscript{77}

Moreover, this form of HLE will not necessarily result in compassion even if the sympathizer \textit{would} suffer in the other’s shoes: there is empirical evidence to suggest that this form of HLE can result in \textit{either} compassion \textit{or} personal distress in response to imagining our own suffering. (Batson, et al., 1997) And which one the imaginer ends up feeling can be influenced by his disposition and character. (Eisenberg & Eggum, 2009) For example, those who are more disposed to feel negative emotions are more likely to feel personal distress rather than compassion as a result of in-his-shoes-imagining. (\textit{Ibid.})\textsuperscript{78}

\begin{footnotes}
\textsuperscript{77} Cf. Griswold (1999, p. 94).
\textsuperscript{78} I return to these issues, and examine their implications in more detail, in Chapters 4, 5 and 6.
\end{footnotes}
Smith could rejoinder here by reminding us that he employs another form of HLE, imagining being the other person, and that it is this form of HLE that causes compassion. The imaginative process that results in sympathy as described so far is in-his-shoes-imagining. As well as this form of HLE, Smith also uses another form that I described in Chapter 1 (§1.3) as imagining being the other person. He mentions this in a section towards the end of TMS.

When I condole with you for the loss of your only son, in order to enter into your grief I do not consider what I, a person of such a character and profession, should suffer, if I had a son, and if that son was unfortunately to die: but I consider what I should suffer if I was really you, and I not only change circumstances with you, but I change persons and characters. (*TMS*, p. 317, VII.iii.1.4)

It is by imagining being the other person, Smith claims, that a man can sympathize with a woman who is in labour even though he cannot possibly ‘conceive himself as suffering her pains in his own proper person and character.’ (*TMS*, p. 317, VII.iii.1.4) It is rather by imagining being her that he comes to sympathize. Smith offers this example to show that sympathy is not a selfish principle, as he explains in the following quotation.

[Sympathy] is not, therefore, in the least selfish. How can that be regarded as a selfish passion, which does not arise even from the imagination of any thing that has befallen, or that relates to myself, in my own proper person and character, but which is entirely occupied about what relates to you? A man may sympathize with a woman in child-bed; though it is impossible that he should conceive himself as suffering her pains in his own proper person and character. (*TMS*, p. 317, VII.iii.1.4)

Perhaps Smith could insist that it is this other type of HLE that causes compassion. Using it would make it less likely that our compassion depended on our feeling misery in the other’s situation, and would not therefore be subject to the foregoing objections. An additional advantage of
this form of HLE is that it is less likely to lead to personal distress than in-his-shoes-imagining. (Batson, et al., 1997) However, the invocation of this other form of HLE seems somewhat *ad hoc*. It occurs during a critique of moral systems ‘which deduce the Principle of Approbation from Self-love.’ (*TMS*, p. 315, VII.iii.1.1) During his critique, Smith is distinguishing his view from theories put forward by Hobbes, Puffendorff and Mandeville, according to whom the concern we have for the welfare of others derives from self-interest rather than sympathy. (*TMS*, p. 316, VII.iii.1.3) The main form of HLE employed throughout *TMS*, however, is in-his-shoes-imagining. And this form of HLE is introduced when he is setting out his theory of sympathy and how it operates. Smith is explicit at the beginning of *TMS* when he introduces sympathy that it is *this* form that is necessary for compassion, as the following quotation demonstrates.

> The compassion of the spectator must arise altogether from the consideration of what he himself would feel if he was reduced to the same unhappy situation. (*TMS*, p. 12, I.i.1.11)

Furthermore, it is in-his-shoes-imagining that Smith relies on in his moral theory and his account of virtue, and not imagining being the other person. We ought therefore to consider the former to be the process necessary for sympathy and compassion. If this is true, then we can conclude, albeit tentatively given the ambiguities present in the text, that HLE is not sufficient for compassion because the disposition and character of the spectator will affect the outcome of the HLE.

### 3.3.3 Summary

I have argued in this section that Smith faces aspects (1) and (2) of the motivation objection in relation to his claim that HLE is sufficient for compassion. I argued that, given Smith’s claims that HLE is the only way we can know how the other feels, and that compassion arises as the result of
another person’s misery and suffering, it is plausible to assume that attributing suffering to the other is necessary for compassion. I argued that HLE processes are unreliable, frequently giving rise to incorrect attributions of mental states to others because simulators are subject to quarantine failure. If compassion is dependent on making a correct attribution of suffering to the other person, then HLE cannot be sufficient for compassion. 

I then argued that Smith could accept this: given his examples of illusive sympathy, perhaps an attribution of suffering is not necessary for compassion. However, I proceeded to argue that Smith faces aspect (2) of the motivation objection, because it is not the case that we would always feel the same as the other feels in his situation, something upon which feeling compassion does depend. I argued that this form of HLE will not necessarily result in compassion, even if the sympathizer would suffer in the other’s shoes, because our disposition and character can determine whether or not we feel personal distress or compassion as a result of engaging in this imaginative act. Smith could respond that it is imagining being the other person that compassion in fact depends on, but I concluded that the weight of evidence is in favour of compassion being caused by in-his-shoes-imagining, and it is this that forms the basis of his account of moral evaluation. We should therefore consider it to be the process necessary for sympathy and compassion, in which case aspect (2) still applies: HLE is not sufficient for compassion because our motivations can influence whether or not we feel compassion as a result of HLE.

3.4 Moral Evaluation

In this section I apply aspect (2) of the motivation objection to Smith’s theory of moral judgement. In the previous section, the claim was that HLE is insufficient for compassion because our dispositions and characters can affect the outcome of the HLE. In this section, however, as was the case in Chapter 2 (§2.4), the claim is that HLE is not sufficient for moral judgement.
I argue that the outcome of our HLE can be affected by our disposition and character. On Smith’s view, moral judgement requires that the spectator put himself in the other’s shoes and imagine if he would be motivated to act as the actor does. If he would be, he will sympathize, and approve. However, I argue that HLE is not sufficient for moral judgement because we tend to think we would feel differently in others’ shoes than we perhaps really would. Our moral judgements of others can therefore be affected by our subjective sentiments, for we can sympathize and approve (or fail to sympathize and disapprove) of others based on what we falsely imagine we would feel in their shoes. However, I argue that Smith can invoke the device of the impartial spectator in order to ensure that our moral judgements are not affected by our self-evaluations and are thus impartial. To make moral judgements of others, he argues, we must first of all adopt the perspective of an impartial spectator and make judgements of self-approval. If the impartial spectator would approve of our conduct, then we can use this as a standard against which to measure the conduct of others. I then argue that this requires spectators to have a reason (a motive) to adopt the perspective of the impartial spectator and judge themselves according to his prescriptions. It is always the perspective of the impartial spectator that we adopt, for his approval is most important and is what we appeal to in the face of dispute and disagreement with others. Smith could argue that we seek the sympathy and approval of the impartial spectator because it is pleasurable. However, agents could surely obtain pleasure from the sympathy and approval of those who do not represent the impartial spectator. And we can surely self-approve without adopting the perspective of the impartial spectator, if for example we act according to our own values, and then sympathize with ourselves. We might, in other words, evaluate ourselves according to criteria that are not what the impartial spectator would approve of, and yet still feel pleasure. So Smith must provide another
reason why we are motivated to HLE with, and sympathize with, the impartial spectator. Why should we care whether or not he approves of us? Smith rejects the idea that we seek sympathy with the impartial spectator because of self-love, or because of the love of mankind, so he cannot appeal to these as reasons why we seek the sympathy of the impartial spectator. It seems that, on Smith’s view, we seek sympathy with the impartial spectator because we want to be virtuous.

3.4.1 Introduction

HLE is, on Smith’s view, necessary for moral judgement. It is by putting ourselves in the other person’s shoes and feeling a sympathetic sentiment that we approve of him. Disapproval is a matter of failing to sympathize as a result of HLE. As was the case for Hume, approval and disapproval is a matter of feeling moral sentiments of approbation or disapprobation. In §3.2 above we saw how HLE and sympathy can cause approbation. HLE causes sympathy, and ‘our approbation arises from sympathy.’ (TMS, p. 325, VII.iii.3.14) Figure 9 below is an adaptation of Figure 8 (§3.2.4), modified to include disapprobation. A spectator comes to feel approbation as follows. He puts himself in the actor’s shoes. If he comes to feel as the actor feels, then he sympathizes. If he sympathizes, he feels approbation. If he fails to sympathize, then he feels disapprobation.
Smith claims that approbation requires sympathy, which is having a correspondence of sentiments with another person, and perceiving this correspondence of sentiments between ourselves and that another person. The sentiments of approbation and disapprobation can feel slightly different according to what it is we are approving or disapproving of. (TMS, pp. 67, II.i.intro.1; 324-327, VII.iii.3.13-17) The sympathetic feeling can be pleasurable or painful, but the sentiment of approbation is always pleasurable. Disapprobation is always painful. In the following quotation, a response to an objection made by Hume, Smith outlines exactly what the sentiment of approbation is. On Hume’s view, approval is pleasurable but arises as a result of sympathy with the pleasure felt by a person on the

---

79 This is similar to Hume’s view: as we saw in the previous chapter (§2.4.1), he argued that there can be more than one moral sentiment.
receiving end of a virtuous action. But for Smith, sympathy can be pleasurable or painful. Yet the sentiment of approbation is, on his view, always pleasurable. Hume demanded to know how, if the sentiment of approbation was always pleasurable, Smith could claim that we can sympathize with passions that are both painful and pleasurable, that is, with ‘any passion whatever’. If it were really the case that approbation could be pleasurable even though sympathy is painful, then surely, Hume claimed, ‘an Hospital woud [sic] be a more entertaining Place than a Ball.’ (TMS, p. 46, editors’ footnote 2) In response, Smith clarified what he meant by approbation:

It has been objected to me that as I found the sentiment of approbation, which is always agreeable, upon sympathy, it is inconsistent with my system to admit any disagreeable sympathy. I answer, that in the sentiment of approbation there are two things to be taken notice of; first, the sympathetic passion of the spectator; and, secondly, the emotion which arises from his observing the perfect coincidence between this sympathetic passion in himself, and the original passion in the person principally concerned. This last emotion, in which the sentiment of approbation properly consists, is always agreeable and delightful. The other may either be agreeable or disagreeable, according to the nature of the original passion, whose features it must always, in some measure, retain. (TMS, p. 46, I.iii.1.9)

According to this quotation, a ‘perfect’ correspondence between the sentiments of the other person and our own is required for approval. And Smith claims elsewhere that approval depends on perfect correspondence, as in the following quotation, for example.

To approve of the passions of another, therefore, as suitable to their objects, is the same thing as to observe that we entirely sympathize with them; and not to approve of them as such, is the same thing as to observe that we do not entirely sympathize with them. (TMS, p. 16, I.i.3.1, my italics)
Approbation necessarily follows from sympathy, Smith claims. HLE is sufficient for moral approval because it is sufficient for sympathy, and approval necessarily follows from this. That sympathy entails approval is evident from the following quotation.⁸⁰

When the original passions of the person principally concerned are in perfect concord with the sympathetic emotions of the spectator, they necessarily appear to this last just and proper, and suitable to their objects; and, on the contrary, when, upon bringing the case home to himself, he finds that they do not coincide with what he feels, they necessarily appear to him unjust and improper, and unsuitable to the causes which excite them. (TMS, p. 16, I.i.3.1, my italics)

3.4.1.1 Propriety and Impropriety

On Smith’s view, our actions can be judged according to their propriety and impropriety and according to their merit and demerit. (TMS, pp. 18, I.i.3.6; 67, II.i.intro.1) I discuss judgements of merit and demerit in §3.4.1.2. Judgements of propriety and impropriety are made as a result of in-his-shoes-imagining. If the imaginer sympathizes, that is, if he comes to feel what the other feels, he sympathizes and approves. What makes the action or sentiment of another person proper or improper is its suitableness or unsuitableness to the cause that gave rise to it. If S would feel the same sentiment in response to the same cause, then he will sympathize if, when he puts himself in A’s shoes, he comes to feel the same sentiment in response to the same cause. He will then approve of A’s feeling as proper or appropriate. It is not only feeling the same sentiment as the other person in the same circumstances and given the same causes that determines S’s judgement of propriety, however. It is also the pitch or level of A’s sentiment. S will approve of A if he, S, would feel what A is feeling in his situation, at the same pitch or level.

⁸⁰ ‘Perfect correspondence’ does not mean ‘exact replica’: the feeling or emotion that a spectator feels need only be similar to that of the actor to count as sympathy. See §3.2.4.4 above.
He who laughs at the same joke, and laughs along with me, cannot well deny the propriety of my laughter. On the contrary, the person who, upon these different occasions, either feels no such emotion as that which I feel, or feels none that bears any proportion to mine, cannot avoid disapproving my sentiments on account of their dissonance with his own. (TMS, p. 16, I.i.3.1)

What determines the proper level of each passion in relation to its cause is the extent to which we can sympathize with it, and each passion has a specific ‘mediocrity’, a level at which it is most easily sympathized with, which determines the propriety of its expression. As we saw earlier (§3.3.2), we are more disposed, for example, to sympathize with passions that arise from the imagination, and much less to sympathize with those originating in the body. Expression of bodily passions can never be fully sympathized with, and as a result can never be considered proper. As Smith puts it:

[ ...] if we consider all the different passions of human nature, we shall find that they are regarded as decent, or indecent, just in proportion as mankind are more or less disposed to sympathize with them. (TMS, p. 27, I.ii.intro.1)

3.4.1.2 Merit and Demerit

Other people’s actions and conduct can be considered meritorious or demeritorious, as well as proper or improper. (TMS, pp. 18, I.i.3.6; 67, II.i.intro.1) Merit and demerit are ‘the qualities of deserving reward, and of deserving punishment.’ (TMS, p. 67, II.i.intro.1) We apportion the quality of merit to an action if the motive behind it is aimed at the good, or tends to produce a good effect. (TMS, p. 67, II.i.intro.2) *Mutatis mutandis* for demerit. As Smith puts it:

...upon the beneficial or hurtful effects which the affection proposes or tends to produce, depends the merit or demerit, the good or ill desert of the action to which it gives occasion. (TMS, p. 67, II.i.intro.2)
Whether an action deserves reward or punishment is determined by whether it causes proper feelings of gratitude (for reward) or resentment (for punishment) in the recipient. On Smith’s view, gratitude ‘immediately and directly prompts us to reward, or to do good to another.’ (TMS, p. 67, II.i.1.1) Resentment, on the other hand, is the sentiment that ‘immediately and directly prompts us to punish, or to inflict evil on another.’ (TMS, p. 68, II.i.1.1) It is helpful to understand approval (and disapproval) of actions for their merit (or demerit) in terms of a three-way relationship between S (spectator), A (actor) and R (recipient). (Broadie, 2006, p. 178) If A’s action causes proper gratitude in R, it will immediately prompt S to want to reward A. If A’s action causes proper resentment in R, it will immediately prompt S to want to punish A. So actions that cause proper gratitude deserve reward, and are therefore judged meritorious, and those which cause proper resentment deserve punishment, and are therefore judged demeritorious.

S’s moral sentiments are different in the case of merit and demerit to his sentiments in the case of propriety and impropriety. (TMS, p. 67, II.i.intro.1) The feeling of disapprobation involved in judgements of demerit is a ‘compound’ sentiment comprising ‘two distinct emotions; a direct antipathy to the sentiments of the agent, and an indirect sympathy with the resentment of the sufferer.’ (TMS, p. 75, II.i.5.5) In the case of approbation, S’s feeling is a compound sentiment consisting of direct sympathy with A’s motives and indirect sympathy with R’s gratitude.

3.4.2 Aspect (2): Moral Evaluation

Aspect (2) of the motivation objection is the claim that empathy is not sufficient for compassion because our motives can influence whether or not we feel compassion as a result of empathizing. In this section I apply aspect (2) to Smith’s claim that HLE is sufficient for moral judgement, or rather the
moral sentiments of approbation and disapprobation that follow on from sympathy, which is the outcome of HLE. More specifically, HLE is not sufficient for impartial moral judgement because our disposition and character can affect the outcome of the HLE, viz. the sympathy. For whether or not we will sympathize and approve, or fail to sympathize and disapprove, can be affected by subjective—and perhaps even false—judgements about how we imagine we would feel in the other’s situation. So our moral judgements of others will be partial in the sense that they will be made according to how we ourselves—sometimes mistakenly—think we will feel.

Whether or not S feels sympathy as a result of HLE, and therefore whether or not he feels approbation towards A, depends on whether S would himself feel the same passion to the same degree in A’s situation. S’s motives—his disposition and character—will affect whether he would feel the same passion, to the same level, and therefore whether or not he will approve of A. S might well imagine that, in A’s shoes, he would feel either the same passion at a different level, or a different passion altogether. We saw some possible effects of egocentric bias in the discussion of simulation (§3.3.1) in relation to third-person emotion attributions. Egocentric bias also affects our attributions of mental states to ourselves. Psychological studies show that people’s self-evaluations are subject to significant biases, biases that are ‘inflicted on us by our wants, needs and desires.’ (Moskowitz, 2005, p. 312) We often make false self-evaluations, such as tending to see ourselves as morally superior to others. This affects how we imagine we would feel in another person’s situation. We tend to believe ourselves to be more moral than others—more generous, kind and so forth—and we tend to underestimate our negative virtues, believing of ourselves that we are less selfish, greedy, and so on, than others. (Epley & Dunning, 2000, p. 861)
There is also evidence of this so-called ‘moral confabulation’ in studies exploring how we think we would act in certain situations. According to Smith, an action can be evaluated according to the motive behind it and the effect it either has or tends to have, or aims to have. (TMS, p. 18, I.i.3.5) S will approve of A’s action as proper if he sympathizes with the motive or sentiment behind it as proper to its cause. For example, if A laughs at a joke that S would find equally amusing, then S will approve of A’s laughter as proper and appropriate to its cause, \textit{viz.} the funniness of the joke. (TMS, p. 16, I.i.3.1) S will approve of A’s action as meritorious if the motive behind it ‘aims at, or tends to produce’ positive effects. (TMS, p. 16, I.i.18.7) Studies have shown that we have a tendency to make false evaluations of our own motives. We have a tendency to believe, for example, that we would not give people electric shocks if we were prison guards in Stanley Milgram’s well-known experiment, and that we would always give up our seat to a pregnant woman on a bus. (Haidt & Kesebir, 2010, p. 811) Why we are motivated to make these false self-evaluations varies. It has been argued that we make them out of a need to maintain a positive sense of self (Moskowitz, 2005, p. 312), and out of a desire to increase our own self-esteem. (Sedikides & Strube, 1997)

I have argued that our self-attributions or evaluations can be affected by our motives. Our conclusions about how we think we would feel in another person’s shoes are necessary for sympathy, because sympathy is felt as a result of imagining how we would feel in their shoes. Approval necessarily follows from sympathy. So it is reasonable to conclude that, if we sympathize with another person based on how we wrongly imagine we will feel in their shoes, we will make biased or partial or even false moral judgements of others. Earlier (§3.2.3) we considered the example of A, the angry man, a father whose son had been assaulted by R. Imagine that A
takes the law into his own hands, meting out his own brand of vigilante justice on R. If S wrongly believes he would not be motivated to act as A did, he will not sympathize with, and will disapprove of, A’s anger and subsequent act of retribution. He may, however, sympathize with R’s resentment. In this case he will feel the compound sympathy necessary for judgements of demerit, comprised of an antipathy with A’s motive and sympathy with R’s resentment. As we saw earlier (§3.4.1.2), if we judge an action to be demeritorious then we consider it deserving of punishment. It is reasonable to assume, therefore, that S would consider A worthy of punishment. But what if A has actually acted rightly? S will disapprove of him anyway, and will consider his action worthy of punishment. Smith’s theory seems to make approval and disapproval entirely subjective: our approval or disapproval of the other person’s motives is entirely dependent on our own, whether that is what our own actually would be, or what we falsely imagine they will be. But surely we want to be able to say that we approve based on someone’s meeting some objective standard, which is something Smith does not seem to be able to say.

In response to the objection that tying approval to sympathy renders moral judgement non-objective, Smith would presumably invoke his notion of the impartial spectator. We adopt the point of view of the impartial spectator, Smith claims, in order to make moral judgements of ourselves. In TMS Smith explains how we make our moral judgements of our own actions and sentiments after having explained how we make judgements of others’ actions and sentiments. In self-judgement, we ‘endeavour to examine our own conduct as we imagine any other fair and impartial spectator would examine it.’ (TMS, p. 110, III.i.3) If, Smith argues, ‘upon placing ourselves in his situation, we thoroughly enter into all the passions and motives which influenced it’, we will approve of our own conduct. (Ibid.) In other words, we
will approve of our own conduct if, when we put ourselves in the impartial spectator’s shoes, we sympathize with ourselves. Conversely, if we put ourselves in his shoes and fail to sympathize with ourselves, we will disapprove of our own conduct. The examination of our own conduct is no mean feat. It involves a complicated exercise of the imagination, which Smith describes in the following excerpt from *TMS*.

> When I endeavour to examine my own conduct, when I endeavour to pass sentence upon it, and either to approve or condemn it, it is evident that, in all such cases, I divide myself, as it were, into two persons; and that I, the examiner and the judge, represent a different character from that other I, the person whose conduct is examined into and judged of. The first is the spectator, whose sentiments with regard to my own conduct I endeavour to enter into, by placing myself in his situation, and by considering how it would appear to me, when seen from that particular point of view. The second is the agent, the person whom I properly call myself, and of whose conduct, under the character of a spectator, I was endeavouring to form some opinion. The first is the judge; the second the person judged of. But that the judge should, in every respect, be the same with the person judged of, is as impossible, as that the cause should, in every respect, be the same with the effect. (*TMS*, p. 113, III.1.6)

Once we have made a judgement of our own conduct from the impartial spectator’s perspective, and have concluded that he would sympathize with us, we can utilise this standard against which to judge the sentiments and conduct of others. If we then adopt this standard when judging the other, we can denominate *their* actions right or wrong according to whether we, having adopted this standard, would sympathize with *them*. Smith is clear that judgements of self-approbation are a necessary prerequisite of judgements of others, and that it is the impartial spectator’s point of view that is to be appealed to when we make judgements of self-approbation. Indeed, it is the perspective of the impartial spectator that ultimately
represents our consciences. (TMS, p. 134, III.3.1) For example, take the following quotation.

But though man has, in this manner, been rendered the immediate judge of mankind, he has been rendered so only in the first instance; and an appeal lies from his sentence to a much higher tribunal, to the tribunal of their own consciences, to that of the supposed impartial and well-informed spectator, to that of the man within the breast, the great judge and arbiter of their conduct. (TMS, p. 131, III.2.33)

However, we can now see how Smith faces a problem similar to one that arose in Chapter 2. In Chapter 2 (§2.4.3) we saw that, on Hume’s view, we make our moral judgements from a common point of view. We also make judgements from our own, situated points of view. As a result, we feel two sentiments, one from the situated point of view and one from the common point of view. We must have a motive to privilege the sentiments we feel from the common point of view over our own. In Smith’s case, things are slightly different. He claims that we must adopt the perspective of the impartial spectator and judge ourselves from that vantage point in order to then make impartial moral judgements both of ourselves and of others. We have to sympathize with the impartial spectator, in other words. However, as was the case with Hume, it looks as though the spectator must have a reason to adopt the perspective of the impartial spectator so that he can sympathize. This will require being motivated to put ourselves in his shoes, for it is by this means that we come to sympathize with, and therefore approve of, ourselves.

The motivation for adopting the point of view of the impartial spectator could be pursuit of the pleasure of mutual sympathy. In the second chapter of TMS, right after having introduced the notion of sympathy, Smith claims that ‘nothing pleases us more than to observe in other men a fellow-feeling
with all the emotions of our own breast; nor are we ever so much shocked as by the appearance of the contrary. ‘(TMS, p. 13, I.i.2.1) We like it when others sympathize with us, and are ‘mortified’ when they do not. (Ibid.) And we like to sympathize with others, and find it painful when we cannot do so. (TMS, p. 15, I.i.2.4-6)

A man is mortified when, after having endeavoured to divert the company, he looks round and sees that nobody laughs at his jests but himself. On the contrary, the mirth of the company is highly agreeable to him, and he regards this correspondence of their sentiments with his own as the greatest applause. (TMS, p. 13, I.i.2.1)

However, given that spectators can obtain pleasure from sympathizing with anyone, not just the impartial spectator (their conscience), what makes the sympathy of the impartial spectator so special? Why, in other words, would spectators seek pleasure from sympathy with and approval of the impartial spectator, when they can surely obtain pleasure from sympathy with, and approval of, other people? Moreover, why can we not seek pleasure from sympathy with, and approval of, ourselves? It is not satisfactory for Smith to reply that we would be motivated to HLE with, and seek sympathy with, the impartial spectator because of the pleasure we would garner from it, given that we can derive pleasure from sympathy with other spectators. Smith must provide us with a better reason why spectators would adopt the point of view of the impartial spectator and seek sympathy with him.

Smith seems to provide one possible motive for seeking the approval of the impartial spectator during his discussion of conscience: that it is because we love to be considered virtuous. We have seen that, on Smith’s view, our consciences represent the impartial spectator. Conscience is a kind of third-person perspective; it is the means by which we can determine what the
right thing to do is, independently of our own interests and the interests of others. Conscience is not our own perspective or that of other men, but rather ‘a third person, who has no particular connexion with either, and who judges with impartiality between us.’ (TMS, p. 135, III.3.3) Smith argues that we adopt this third-person perspective because it is the means by which we can make impartial judgements about what we ought to do that are not subject to influence either by our own interests or by those of the other person. *(Ibid.)* We know what we ought to do—what conscience dictates—by consulting the impartial spectator. But just because we know what we ought to do, Smith acknowledges, this does not mean that we are automatically motivated to do it. Smith asks us to suppose that an earthquake in China has occurred, causing millions of deaths. ‘Let us consider,’ Smith says, ‘how a man of humanity in Europe, who had no sort of connexion with that part of the world, would be affected upon receiving intelligence of this dreadful calamity.’ *(TMS, p. 136, III.3.4)* After having lamented the fate of those unfortunate people, Smith claims, the man would in all probability go on with his daily life as if nothing had happened, his own ‘paltry misfortune’ preoccupying him more than the deaths of millions of his fellow men. *(TMS, p. 136, III.3.4)* But we are never so inhumane as to actively sacrifice the lives of millions of others to prevent our own misfortune. *(TMS, p. 136, III.3.4)* So what accounts for this difference? As Smith puts it:

When our passive feelings are almost always so sordid and so selfish, how comes it that our active principles should often be so generous and so noble? When we are always so much more deeply affected by whatever concerns ourselves, than by whatever concerns other men; what is it which prompts the generous, upon all occasions, and the mean upon many, to sacrifice their own interests to the greater interests of others? *(TMS, p. 137, III.3.5)*

It is not self-love that makes us do what we know we ought to do, Smith argues. *(TMS, p. 136, III.3.4)* Neither is it love of mankind or love of our
neighbour. (Ibid.) So what does prompt us ‘to the practice of those divine virtues’? (Ibid.) Smith’s answer is as follows.

It is a stronger love, a more powerful affection, which generally takes place upon such occasions; the love of what is honourable and noble, of the grandeur, and dignity, and superiority of our own characters. (TMS, p. 136, III.3.4)

It seems from this quotation that we act according to conscience, which is acting in such ways as the impartial spectator would approve of, because of the love of having a virtuous character. It seems reasonable to conclude from this that we seek sympathy with, and the approval of, the impartial spectator because we value virtue.

3.4.3 Summary

In this section I applied aspect (2) of the motivation objection to Smith’s theory of moral judgement. I argued that sympathy, the outcome of our HLE, can be affected by our disposition and character because we tend to imagine that we would feel differently in others’ shoes than we perhaps really would, for several reasons. Our moral judgements of others can therefore be affected by our motives, because we can sympathize with and approve (or fail to sympathize with, and disapprove) of others based on what we falsely imagine we would feel in their shoes. I argued that Smith might invoke the device of the impartial spectator in order to ensure the impartiality of our moral judgements. I then argued that spectators must have a reason (a motive) to adopt the perspective of the impartial spectator and judge themselves according to his prescriptions, and one such motive to seek the sympathy and approval of the impartial spectator is because it is pleasurable. However, I then argued that spectators can feel pleasure from sympathy and approval from perspectives other than that of the impartial spectator, which means it behoves Smith to provide a further reason why we would be motivated to seek his sympathy and approval. I argued that Smith
rejects the idea that we seek sympathy with the impartial spectator because of self-love or benevolence, and concluded that, on Smith’s view, it seems that we seek sympathy with the impartial spectator because we love virtue.

3.5 Virtue

In this section I apply aspect (3) of the motivation objection to Smith’s explanation of virtue in terms of sympathy. Aspect (3) is the claim that empathy is neutral about motive so using it to define or explain certain aspects of morality in a non-circular way is problematic. I argue that Smith’s explanation of virtue in terms of sympathy is circular because it presupposes what it is trying to explain, namely virtue.

The virtuous person is both a spectator of the actions of others, and an actor himself. Two sets of virtues are, Smith claims, reflected by these two personae of the agent. As an actor, being virtuous is being disposed to make the effort to subdue or bring down our emotions ‘to what the spectator can go along with’. (*TMS*, p. 23, I.i.5.1) As a spectator, being virtuous is being disposed to make the effort to ‘enter into the sentiments of the person principally concerned,’ the actor. (*Ibid.*) As Smith puts it:

> Upon these two different efforts, upon that of the spectator to enter into the sentiments of the person principally concerned, and upon that of the person principally concerned, to bring down his emotions to what the spectator can go along with, are founded two different sets of virtues. The soft, the gentle, the amiable virtues, the virtues of candid condescension and indulgent humanity, are founded upon the one: the great, the awful and respectable, the virtues of self-denial, of self-government, of that command of the passions which subjects all the movements of our nature to what our own dignity and honour, and the propriety of our own conduct require, take their origin from the other. (*TMS*, p. 23, I.i.5.1)
The amiable virtues fall under the general umbrella virtue of ‘sensibility’, the awful virtues under the umbrella of ‘self-command’. On Smith’s view, the perfectly virtuous agent combines these two sets of virtues: he ‘joins, to the most perfect command of his own original and selfish feelings, the most exquisite sensibility both to the original and sympathetic feelings of others.’ (TMS, p. 152, III.3.35) According to Smith, then, perfect virtue involves being fully sensitive to the original and the sympathetic feelings of others, and having total command of our own feelings so that we can adjust them accordingly.

Sensibility and self-command are chief among the virtues because they are the qualities necessary to ensure the sympathy and approval of the impartial spectator. They ensure that our sentiments and conduct will always be at the pitch the impartial spectator can most fully sympathize with, and therefore approve of. With this picture of what virtue is in mind, we can schematise what the virtuous person does as follows.

Step 1: Puts himself in the impartial spectator’s shoes
Step 2: Judges how the impartial spectator feels
Step 3: Adjusts his sentiments to match those of the impartial spectator

Being virtuous requires an exercise of the imagination. As we saw in section §3.4.2, when we make judgements of our own conduct, we divide ourselves into two persons: the impartial spectator—the judge—and ourselves, the person being judged of. We put ourselves in the impartial spectator’s shoes and examine our own conduct through his eyes. (TMS, p. 113, III.1.6) This is Step 1. We then determine whether or not he would sympathize with us, which is Step 2. If we judge that he would do as we do or feel as we feel,
then we judge that he will sympathize with and approve of us. However, if we find that he would not sympathize with us, then we know that he will disapprove of us. In order to ensure that his sentiments are in line with what the impartial spectator would approve of, the virtuous person will adjust them so the impartial spectator can sympathize and approve. This final stage, Step 3, is virtue.

We are now in a position to see how Smith faces aspect (3). On his view, as we have just seen, being virtuous is explained in terms of making the effort to adjust our sentiments at Step 3 so the impartial spectator will approve of us. In order to reach Step 3, an agent must first of all engage in Step 1. That is, he must adopt the perspective of the impartial spectator. However, being concerned to adopt the perspective of an impartial spectator would surely require virtue. For it is not clear what other reason an agent would have to put himself in the impartial spectator’s shoes when he could put himself in the shoes of a spectator whose perspective does not represent the objective moral standard exemplified by the impartial spectator. If putting oneself in the impartial spectator’s shoes requires virtue, then Smith’s explanation of virtue in terms of sympathy will be circular, because if being virtuous is explained as being disposed to obtain the sympathy of the impartial spectator by adjusting our sentiments accordingly, and adopting the perspective of the impartial spectator in the first place requires that we are virtuous, then the explanation of virtue in terms of sympathizing presupposes what it is trying to explain, namely virtue.

Smith could respond that adopting the perspective of the impartial spectator does not require or presuppose virtue because it is simply habitual: he does claim that we learn to put ourselves in the impartial spectator’s shoes ‘so easily and so readily, that we are scarce sensible that we do it.’ (TMS, p. 136,
III.3.3) But in order to become a habit, adopting the perspective of the impartial spectator must be practiced. As Smith himself says of ‘the man of real constancy and firmness’:

[… with the eyes of this great inmate he has always been accustomed to regard whatever relates to himself. This habit has become perfectly familiar to him. He has been in the constant practice [...] of modelling [...] his inward sentiments and feelings, according to those of this awful and respectable judge. (TMS, p. 147, III.3.25)

Adopting the perspective of the impartial spectator is not something we learn to do without practice, and being motivated to practice, it is reasonable to assume, requires that we are aiming at virtue. For why else would we practice adopting the impartial spectator’s point of view so it becomes habitual? We are back to the circle: to be formed into a habit, putting ourselves in the impartial spectator’s shoes requires practice, and practicing requires aiming at virtue. We can conclude that aspect (3) of the motivation objection applies to Smith’s explanation of virtue in terms of sympathizing with the impartial spectator, because sympathizing with the impartial spectator presupposes virtue.

### 3.6 Conclusion

This chapter introduced the second of the three constitutive views I consider in the thesis. According to Smith, sympathy is foundational for morality because it serves as a moral standard against which we can measure our own conduct and the conduct of others. It is necessary for moral judgement and for compassion. The process that I have been calling HLE or in-his-shoes-imagining also plays a foundational role in morality insofar as it is necessary for producing sympathy. It is by putting ourselves in the other’s shoes that we sympathize with them. Sympathy also serves to explain what being
virtuous consists in: being virtuous is seeking sympathy with the impartial spectator.

In this chapter I argued that Smith faces all three aspects of the motivation objection: reliability, compassion and circularity. In section 3.2 I introduced Smith’s concept of sympathy, arguing that it is the outcome of a HLE process, not the process itself. I also argued that Smithean sympathy can only be produced by such a process; LLE processes are insufficient for sympathy. I argued that sympathy can give rise to compassion and the moral sentiment of approbation. I defended the view that Smith is a proto-simulationist against a contemporary interpretation of Smith’s sympathy offered by Bence Nanay.

In section 3.3 I applied aspects (1) and (2) of the motivation objection to Smith’s claim that HLE is sufficient for compassion. I argued that Smith faces aspect (1) because HLE processes are unreliable, frequently giving rise to incorrect attributions of mental states to others. I also argued that compassion is dependent on spectators making correct attributions of suffering to the other person. If this is true, then HLE cannot be sufficient for compassion, given (1). However, I argued that Smith could accept this, for there are cases of sympathy where the object of the sympathy is not suffering. So perhaps correct attributions are unnecessary after all. However, I then brought in aspect (2), arguing that this applies even if attribution of suffering is not required for compassion. Compassion is dependent on HLE because it is only by engaging in it that we can sympathize. We only sympathize if we would feel the same in the other person’s shoes. Compassion therefore depends on whether or not we would suffer in the other person’s situation: if we would not, we will not sympathize or feel compassion. I argued that our disposition and character
can determine whether or not HLE results in our feeling compassion, or another outcome, personal distress. Our disposition and character can affect the outcome of our HLE, and it is therefore insufficient for compassion. Smith could respond by invoking the other form of HLE that he employs near the end of *TMS*. However, I concluded that the weight of evidence is in favour of compassion being caused by in-his-shoes-imagining, and that we ought to deem this his considered view because it forms the basis of his account of moral evaluation. And if this is the case, then HLE remains insufficient for compassion because of motive.

In section 3.4, I argued that Smith faces aspect (2) of the motivation objection in relation to moral judgement. He argues that moral judgement requires the spectator to put himself in the actor’s shoes and imagine if he would act as the actor does with the same motives. If he would, he will sympathize, and also approve, because sympathy entails approval. However, I argued that HLE is not sufficient for moral judgement because our dispositions and characters can cause us to falsely imagine how we would feel in the other’s shoes. Our moral judgements of others can therefore be affected by our motives, potentially resulting in incorrect and partial moral judgements. I argued that Smith could employ his device of the impartial spectator to ensure partiality and the elimination of subjective bias. However, we saw that this requires spectators to have a motive to adopt the perspective of the impartial spectator and seek his approval. Smith could argue that we seek the sympathy and approval of the impartial spectator because it is pleasurable. However, I argued that agents could surely obtain pleasure from the sympathy and approval of many others besides the impartial spectator. Smith must provide another reason why we are motivated to HLE with, and sympathize with, the impartial spectator. He rejects self-love and benevolence as motivations for adopting the perspective of the impartial
spectator, instead claiming that we seek sympathy with the impartial spectator because we want his approval. And we want his approval because his approval means we are acting virtuously. We want to act virtuously because we value virtue.

Finally, in section 3.5 I argued that Smith faces aspect (3) of the motivation objection because his explanation of virtue in terms of sympathy looks circular. According to Smith, virtue involves being fully sensitive to the original and the sympathetic feelings of others, and having total command of our own feelings so that we can adjust them accordingly. Sensibility and self-command ensure that our sentiments and conduct will always be at the pitch the impartial spectator can most fully sympathize with, and therefore approve of. In order to establish whether or not his sentiments need adjusting, an agent must first of all adopt the perspective of the impartial spectator. However, I argued that adopting the perspective of an impartial spectator must require virtue, because the impartial spectator represents an objective moral standard that we must choose to adopt. Smith’s attempt to explain virtue as the disposition to sympathize with the impartial spectator is therefore circular, I argued, because sympathizing with an impartial spectator itself requires virtue. I considered one possible response Smith could make to this objection, viz. that putting oneself in the impartial spectator’s shoes is habitual, and as such does not require virtue. However, in order to become a habit, it would have to be practiced, which in turn requires that the agent is aiming at virtue. I therefore concluded that Smith’s explanation of virtue in terms of sympathy is circular, because it presupposes what it purports to explain: virtue.
4 Michael Slote’s Empathy

4.1 Introduction

In the previous two chapters I explained and analysed two sentimentalist moral theories that have a constitutive role for sympathy. According to Hume and Smith, sympathy plays a vital part in producing compassion and facilitating moral judgement. I argued that Hume and Smith face aspects of the motivation objection. Hume faces aspect (2) in relation to both compassion and moral evaluation; discussion of Hume’s view also acquainted us with aspect (3). Smith, I argued, faces all three aspects. In this chapter, I turn to the final constitutive view I am considering in the thesis: the empathy-based moral sentimentalism of Michael Slote. Slote, I argue, faces aspects (2) and (3) of the motivation objection.

The motivation objection, as we have seen, is the general claim that empathy is neutral about motive: we might empathize for morally good motives, and for morally dubious motives, or out of some other kind of motive, for example self-interest, which may or may not coincide with what is morally appropriate. What is more, the emotions or behaviour that can result from the empathizing can be affected by the motive behind it, with the effect that we can have an inappropriate emotional response or act inappropriately as a result of empathy. The motivation objection has three aspects or dimensions: (1) reliability: the fact that it can be affected and influenced by our motivations means that empathy processes are not reliable, as they do not always lead to the right result(s); (2) function: the fact that it can be affected and influenced by our motivations means that empathy is not sufficient for fulfilling certain moral functions; (3) circularity: and empathy is neutral about motive so using it to define or explain certain aspects of morality in a non-circular way is problematic.
Slote (2010a, p. 56) claims that empathy can ground both normative and metaethical sentimentalism. Not only is empathy necessary for morality, on Slote’s view, it is sufficient for many aspects as well, including, for example, justice, rights, moral obligation, and defining the meaning of our moral terms. The main arguments in Slote’s recent book Moral Sentimentalism\(^{81}\) (2010a) are intended to illustrate how a plausible sentimentalist metaethics can buttress the normative theory advocated in a slightly earlier work, The Ethics of Care and Empathy (2007)\(^{82}\) in which he defends an empathy-based normative sentimentalist moral theory.

Slote’s empathy-based normative ethics focuses in particular on empathy’s purported ability to handle key issues in moral philosophy, for example our moral obligations to help others and to refrain from harming them. Slote is a virtue ethicist who allies himself with the ethics of care. Care ethics, on Slote’s view, is actually a form of virtue ethics. (Held, 2006, p. 551) Slote’s normative ethics amounts to the following claims: our actions are virtuous insofar as they are underpinned by a caring motivation; this caring motive is the primary moral virtue, and empathy is essential for producing and sustaining this caring motivation.

Slote’s sentimentalist metaethics also has a fundamental role for empathy. The empathy-based moral criterion he argues for in his normative theory, that an action is right if and only if it expresses (the motive of) empathic concern, can be backed up, he claims, by a purely sentimentalist metaethics. Slote claims that we feel empathic warmth, which he equates with moral approval, in response to actions that express empathic concern, and we feel its opposite, empathic chill (moral disapproval), in response to actions that

\(^{81}\) Henceforth abbreviated to ‘MS’.

\(^{82}\) Henceforth abbreviated to ‘ECE’.
fail to express this motive. These feelings or attitudes of moral approbation and disapprobation do not rely on the making of moral judgements, according to Slote, and serve to fix the reference of our moral terms. ‘Good’ is whatever causes empathy-based approval and ‘bad’ is whatever causes empathy-based disapproval. (MS, p. 57) Given that empathy-based approval is only caused by actions that express empathic concern, empathic concern is what is morally good, and is what we are referring to when we use the term ‘good’. Similarly, given that empathy-based disapproval is only caused by actions that fail to express empathic concern, failing to show this motive in one’s action renders it morally bad.

I will be addressing three of Slote’s claims in this chapter, in sections 4.3, 4.4 and 4.5, and applying an aspect of the motivation objection to each. The chapter is structured as follows. Section 4.2 contains an overview of Slote’s empathy concept. I argue that Slote is unclear about what he means by empathy, but that we can nevertheless interpret him as claiming that empathy mechanisms cause empathic concern, which is, on his view, the fundamental moral motive. Section 4.3 focuses on Slote’s normative ethics, specifically his claim that empathy is sufficient for moral obligation: we can, he claims, base our moral obligations on our empathic reactions. Slote argues firstly that there is a correlation between the strength of our empathic reactions and the strength of our moral obligations, and secondly that the existence of this correlation makes empathy look like it is what is making the difference between what we consider to be a morally right action and what we consider to be an immoral action. Because our moral obligations seem to correlate so well with our empathic reactions, the presence or absence of empathy (empathic concern) can plausibly be what is making the difference between whether we consider one action right and another wrong. So it looks as though we can use empathy (empathic concern) as a moral criterion
against which to measure the rightness and wrongness of actions: actions will be right if they reflect empathic concern, and wrong if they do not.

However, there is a problem with basing morality on empathy that Slote himself anticipates: the capacity for empathy is inherently biased along racial lines. That is, we find it easier to empathize with members of our own race than with members of other races. This means we will feel more empathic concern for them, and will be more motivated to help them, than members of other races. If we based moral obligation on empathy, and empathy is biased in this way, then helping members of our own race before members of other races simply because they are members of our own race will be what we are obliged to do. Not only would this be an instance of our empathic reactions not correlating with what we take our moral obligations to be, it would also be an unwelcome consequence of Slote’s attempt to base moral obligation on empathy.

Slote denies that there is any evidence for empathy’s racial bias, but I present empirical evidence to contradict this. In *ECE*, Slote claims that, even if evidence of any inherent empathic racial bias were to come to light, it would not be a problem for his view, for any ‘predisposition [toward greater empathy with members of one’s own race] might fade over time with the [empathizer’s] increasing cognitive maturity.’ (*ECE*, p. 35) I argue that there are two ways we can interpret Slote’s rejoinder. Firstly, we can interpret him as claiming that being racially biased is immature: it only happens when we are children, and the bias will fade away as we become mature adult empathizers. Against this, however, I argue that the evidence for racial bias is from research into adult empathy, so empathic bias must still persist into adulthood. Secondly, we could interpret him as meaning that it is our empathy that is immature. Perhaps Slote means that as children we can only
empathize with individuals, but as adults we can empathize with groups as well as individuals. In adults, empathy for groups could ‘trump’ or outweigh empathy for individuals in cases where the initial empathic reaction for a member of one’s own race is stronger. Thus empathy for the member of the other race *qua* member of a suffering minority group that is discriminated against will trump empathy for an individual member of one’s own race, rendering empathy’s racial bias unproblematic. If mature adult empathy for groups can trump immature childhood empathy for individuals, then empathy will never lead to justification of preferential helping along racial lines.

However, if this second interpretation is the one Slote endorses, he faces a problem, for it is not the case that our empathic concern for groups always trumps our empathic concern for individuals. I present evidence of the phenomenon of ‘psychic numbing’ (Slovic, 2007), which results in a total failure to empathize with large groups. In cases of psychic numbing, we can only empathize with the large group if we first of all empathize with the individual. Moreover, there are instances of what we can call ‘competing empathies’, when other empathic reactions ‘trump’ empathy with groups. I argue that Slote need not accept the claim that empathy for groups always trumps empathy for individuals, for he could simply argue that it sometimes does, and whether it does or not in the race case is an empirical question. But I argue that if he accepts the importance of empirical evidence in this case, then he must surely also accept the empirical evidence of racial bias. Finally, I argue that it is plausibly false that implicit empathic bias ever trumps empathy for individuals because there does not appear to be any evidence showing this in the empirical literature on empathic bias. I conclude that Slote therefore faces aspect (2) of the motivation objection, for his claim that empathy is sufficient to ground moral obligation is false.
In section 4.4, I apply aspect (2) to Slote’s theory of moral approval and disapproval. Slote claims that empathizing is sufficient to cause us to feel sentiments of approval and disapproval when others display or fail to display empathic concern in their actions. On Slote’s view, the same empathy mechanisms that cause us to feel empathic concern for others when they are suffering are also responsible for causing observers to feel approval and disapproval when they observe empathic concern or its lack manifested in the actions of others. However, I provide evidence to show that empathy can be modulated by our beliefs and prejudices about the object of the empathy, which affects the outcome of our empathizing. I argue that empathizing is not therefore guaranteed to cause the moral sentiments; it is not sufficient for approval and disapproval because the outcomes of empathizing can be affected by our motives.

I consider one possible way Slote could respond to this objection. He could, I argue, reiterate that the empathy process involved in approval and disapproval is a low-level, contagion-like empathy. Because this is involuntary and automatic, he could argue, it is less likely to be subject to modulation by bias and as such could be sufficient for producing the moral sentiments. However, I argue that this response is unsatisfactory, for two reasons. Firstly, the evidence presented shows that, in fact, low-level empathy can be modulated. Secondly, maintaining that only contagion-like empathy is involved would result in Slote’s account of approval being severely limited. Contagion requires that the empathizer and the object of the empathy are both present, so the sentiment of the latter can be ‘picked up’ by the former. If approval and disapproval can only occur when both spectator and actor are present, then approval and disapproval of, for example, people who are not in close proximity to the spectator would be
impossible. I conclude that empathy is not sufficient for the moral sentiments because it can be modulated by our beliefs and prejudices.

In section 4.5 I apply aspect (3) of the motivation objection to Slote’s attempt to define virtue in terms of empathy. I utilise the evidence of modulation to argue that, because the outcome of our empathy can be affected by our prejudices and biases, empathic concern cannot be used to define virtue in a non-circular way. Slote claims that empathic concern is the fundamental moral virtue, and that empathy mechanisms cause empathic concern. But empathy mechanisms will only produce empathic concern if the empathizer is not prejudiced or biased. And not being prejudiced, it is reasonable to assume, is a precondition of virtue. I conclude that Slote’s definition of virtue in terms of empathy is circular, because it presupposes what it purports to define, namely virtue.

4.2 Empathy

In this section I give an overview of what Slote means by empathy. He considers his empathy-based moral sentimentalism to be a direct descendant of Hume’s, but modifies Hume’s theory in order to transform it into what, he claims, is a more plausible and coherent sentimentalism. For example, he employs the empathy concepts advocated by Martin Hoffman (2000) and Daniel Batson (1991) in developmental and social psychology, rather than Hume’s anachronistic sympathy concept. Although he allies his view to Hume’s, Slote is at pains to distinguish his own brand of sentimentalism from Smith’s, which he considers to be a lot less sentimentalist than he would like. (MS, p. 34) Slote does, however, include within his empathy concept the kind of HLE that Smith employs.
Slote builds what he calls a ‘unified’ moral sentimentalism on the concept of empathy, so-called because it encompasses both normative sentimentalism—in the form of an empathy-based ethics of care—and a version of metaethical sentimentalism according to which empathy-based feelings of approval and disapproval form the basis of, and do not presuppose, the making of moral judgements, and fix the reference of our moral terms. Slote’s normative ethics places an agent’s motives, rather than his character, at its heart, although the latter retains moral significance. (Slote, 2001, p. 7) Because his virtue ethics is agent-based rather than agent-focused, it ‘treats the moral or ethical status of actions as entirely derivative from independent and fundamental ethical/aretaic facts (or claims) about the motives, dispositions, or inner life of moral individuals.’ (Slote, 2001, p. 7, Slote’s italics)

On Slote’s view, empathic concern, also referred to as ‘empathic caring’ (MS, p. 52), is the fundamental moral motive: being moral ‘amounts to being (fully) empathically caring with others.’ (Slote, 2010b, p. 131) And being empathically caring towards others is the primary moral virtue. (Held, 2006, p. 551) Traditional care ethics sees the caring relationship as paramount, but for Slote the agent’s caring motive is what is important. (Ibid.) Moreover, Slote argues that this fundamental motive or virtue can underpin other areas of morality, such as moral obligation, as we will see in §4.3. Traditional care ethics, on the other hand, considers aspects of caring other than motive to be equally important. For example, the outcome or consequence of the caring behaviour—what the caring action leads to—is as important to acknowledge, they claim, as the caring motive itself. (Ibid.) And not all care ethicists agree

83 ‘Aretaic’ is from the Greek word for excellence or virtue. It is contrasted by Slote with ‘deontic’, from the Greek word for necessity. (Slote, 2001, p. 4) Agent-focused virtue ethics focuses on inner traits or dispositions of the agent when describing the virtuous individual. Slote’s agent-based ethics considers an agent’s virtue more in terms of his or her motives: an action is right and virtuous if it was done with a virtuous motive.

84 See Chapter 1, §1.2 for an explanation of empathic concern.
with Slote that other areas of morality ought to be subsumed within the notion of caring. Virginia Held (1984), for example, argues that the domain of caring is conceptually distinct from other spheres of morality, such as justice, and should be integrated with caring rather than incorporated under its rubric. Slote, on the other hand, claims that an ethics of care based on empathy can encompass or incorporate all moral domains.

Empathic concern is displayed ‘in the form of kindness, love, benevolence, compassion, humanitarianism, and the like’ (MS, p. 76) and is produced by what Slote refers to as ‘empathy mechanisms’. (Slote, 2010a, p. 128) These mechanisms are based on those put forward by Martin Hoffman (2000). Slote refers the reader to Hoffman’s account of empathy at various points in MS (e.g. MS, p. 30), ECE and elsewhere (e.g. 2010b, p. 132), although he does not go into much detail about what exactly empathizing involves, spending only a few pages per book on the workings of the empathy mechanisms and what exactly he means by ‘empathy’. It is unclear whether Slote is referring to the mechanisms that produce empathic concern, or the motive of empathic concern itself, when he uses the term ‘empathy’. Several commentators have made this point, arguing that Slote’s definition of empathy is somewhat opaque. For example, Angela M. Smith (2010, p. 198) wonders whether empathy is ‘a feeling, an involuntary mechanism whereby feelings are transferred from person to person, or a deliberate imaginative act that we engage in’.85

However, Slote does attempt to define empathy by distinguishing it from sympathy. The former is, he argues, akin to ‘feeling someone’s pain’; the latter is more like ‘feeling for someone who is in pain.’ (MS, p. 15, Slote’s

---

85Baier (2010) and Darwall (2010) make a similar complaint.
Empathy involves ‘having the feelings of another (involuntarily) aroused in ourselves, as when we see another person in pain.’ (MS, p. 15) Slote likens empathy to Hume’s sympathy, as it involves having another person’s feeling ‘infused’ into us by ‘contagion’. (Ibid.) However, it is not sufficiently clear why Slote makes a distinction between empathy and sympathy, given that empathic concern so closely resembles sympathy; after all, it appears in the form of kindness, love, benevolence, compassion, humanitarianism, and the like. Moreover, Batson and Hoffman, whose views Slote claims to endorse, both consider empathic concern (empathic or sympathetic distress, in Hoffman’s case) to be more like sympathy than empathy, as we will now see.

The term ‘empathic concern’ is adopted from the social psychological work on empathy by Daniel Batson (1991). Slote (MS, p. 16) endorses Batson’s empathy-altruism hypothesis, according to which empathy produces and sustains altruistic behaviour for others according to their perceived need. What Batson means by ‘empathy’ in his empathy-altruism hypothesis is empathic concern. For Batson, empathic concern includes emotions like sympathy, compassion and tenderness. All these are what he calls ‘other-oriented’ emotions that are ‘elicited by and congruent with the perceived welfare of another in need.’ (Batson & Eklund, 2007, p. 65) A difference between Batson’s and Slote’s views about empathic concern lies in the fact that Batson separates the feeling or emotion of empathic concern from the motivation to help, and does not define empathic concern as including the motivation to help, for he considers these to be distinct psychological states. (Batson, 2009b) On Slote’s view, however, empathic concern includes a motivation to help another. Indeed, the fact that it does is crucial to Slote’s overall approach: he describes empathic concern as the fundamental moral motive.
According to Hoffman (2000), on whose developmental theory of empathy Slote draws heavily, empathy mechanisms cause what he refers to as empathic distress when we perceive another person in pain or suffering. Empathic distress then gets transformed into four ‘empathic affects’: sympathetic distress (compassion for the victim); empathic anger; empathic feeling of injustice; and guilt over inaction. (Hoffman, 2000, p. 8) Sympathetic distress therefore resembles compassion, which Slote claims is a manifestation of empathic concern or empathy. So again, it looks as though Slote’s notion of empathic concern is very similar to sympathy.

Stephen Darwall (2010, p. 153) points out that what Slote means by empathy differs very little from Francis Hutcheson’s concept of benevolence. Indeed, Slote (2010, p. 126) describes empathy as a ‘folksier’ name for the ‘somewhat formal’ benevolence. So here we see ‘empathy’ described as benevolence, and not the mechanisms that might produce it.

There appears to be confusion surrounding whether Slote’s empathy concept refers to processes that produce empathic affect, or to the outputs of these processes themselves. Furthermore, as we have seen, it is unclear what motivates Slote’s distinction between empathy and sympathy, given that his notion of empathic concern seems to be very similar to sympathy. We can go some considerable way towards clearing up this confusion by observing that, on Slote’s view, empathy ‘lies behind our capacity for moral caring,’ (Slote, 2010b, pp. 126-127) and that ‘empathy is a crucial source and sustainer of altruistic concern or caring about (the well-being of) others.’ (ECE, p. 15) It seems reasonable to read Slote as meaning that empathy ‘lies behind’ empathic concern insofar as it produces it. Or, put another way, empathy is the ‘motivating psychological mechanism’ that produces empathic concern, benevolence, compassion and sympathy. (MS, pp. 5, 79) Given Slote’s
endorsement of Hoffman’s and Batson’s views, both of which consider empathic concern (or sympathetic distress, in Hoffman’s case) to be products of empathy mechanisms, it seems reasonable to ascribe this view to Slote. Moreover, Slote refers to empathy as a ‘causal mechanism’. (MS, p. 14) Empathy mechanisms, we can therefore conclude, cause empathic concern.

Slote argues, following Hoffman, that there are two basic kinds of empathy process: associative empathy and projective empathy. (MS, p. 17) The latter involves deliberately putting oneself in another person’s shoes and the former involves having the feelings of the other person involuntarily aroused in ourselves. (MS, p. 17, footnote 5) The ability to empathize develops along a continuum from birth, culminating in ‘full-blown’, mature adult empathy in adolescence. (MS, pp. 63, 112) Slote does not go into the details of Hoffman’s rich developmental account. Nor does he describe all the stages of empathy’s development, except to say that our capacity for empathy develops along a timeline from newborn reactive cry in babies, which operates ‘via a kind of mimicry and seems like a form of “contagion”’ (MS, p. 17), and which involves babies becoming distressed and crying when they hear other infants and babies crying. As the child gets older and develops more advanced cognitive and linguistic skills, a ‘richer history of personal experience and a fuller sense of the reality of others,’ Slote claims that our empathic reactions are still involuntarily aroused but are ‘mediated’ in the sense that they can be aroused ‘in response to situations or experiences that are not immediately present and are merely heard of, remembered, or read about.’ (MS, p. 17; ECE, p. 14)

As they get older, children develop the ability to ‘projectively’ empathize. That is, they become able to ‘deliberately adopt the point of view of other people and see and feel things from their perspective.’ (MS, p. 17) So as well
as being involuntarily aroused in us, our empathic concern can be aroused as a result of deliberately taking on the perspective of another person. Neither mature associative empathy nor projective empathy involves self-other merging; the empathizer always retains an awareness that he is a separate individual from the target of the empathizing. (MS, p. 17)

Mature empathizing is not limited to feeling what the other feels, Slote claims. Nor is it even dependant on the other feeling pain or suffering. It merely needs to ‘involve feelings or thoughts that are in some sense more “appropriate” to the situation of the person(s) empathized with than to the situation of the person empathizing.’ (MS, p. 17) The empathizer can feel, for example, empathic sadness for a person with terminal cancer who is unaware of his condition, even though he is ‘boisterously’ enjoying himself. (MS, p. 17) This looks like sympathy, and lends further weight to my reading of Slote: the empathizer is feeling sorrow for the sufferer because of his situation. As Slote himself puts it:

In general, as we become more aware of the future or hypothetical results of actions and events in the world, we learn to empathize not just with what people are actually feeling but with what they will feel or what they would feel, if we did certain things or if certain things happened. It seems we even learn to empathize with their (situated) condition and not just with their hypothetical or actual reactions to it. (MS, p. 18)

Slote also claims that as they get older and become aware of the existence of ‘groups or classes of people and the common goals or interests that may unite them,’ children become able to empathize with these groups. (MS, p. 18)

---

86 It is interesting to note the similarity between this example and one offered by Smith that we saw in §3.2.4.2. In Smith’s example, we can sympathize with the man who has lost his reason, even though he himself is unaware of his condition, and even ‘laughs and sings’. (TMS, p. 12, i.i.1.11)
18) Mature adult empathizers are able, Slote claims, to feel empathic concern for individuals and groups.\textsuperscript{87}

4.3 Moral Obligation

4.3.1 Introduction

Slote claims that empathy is a sufficient basis for a moral criterion. In this section I apply aspect (2) of the motivation objection to this claim: I argue that it is false. Aspect (2) is the function aspect: empathy is not sufficient for fulfilling certain functions because it can be affected and influenced by our motivations. Slote faces aspect (2), I will argue, because we are disposed to feel more empathic concern for members of our own race than members of other races, so using empathy as a moral criterion would lead to endorsement and justification of immoral actions such as preferentially helping members of our own race over members of other races. This, I argue, is an unwelcome consequence of Slote’s position.

On Slote’s view, if empathy is employed as the foundation-stone for an ethics of care, it can enable this normative view to rival other ethical theories like Kantianism because it can account for areas of normative ethics seen as out of the normal scope of care ethics. Traditional care ethics, he argues, has been considered unsatisfactory because it cannot account for our caring obligations to strangers, or those outside what Nell Noddings (2010, p. 11) has referred to as an agent’s ‘circle’. Furthermore, Slote argues that traditional care ethics has been melded with other ethical views because it is seen—even by care ethicists like Held (1984)—as incomplete, partly due to the fact that it is typically seen as unable to cope with central issues of deontology, such as our obligations not to harm others. (ECE, p. 11) Slote seeks to provide an ethics of care that can wholeheartedly rival Kantianism.

\textsuperscript{87} This will be important in §4.3.2.
and utilitarianism and he claims that one based on empathy can do this, for it can explain and justify important features of morality that those both within and outside of care ethics have previously considered beyond its scope.

Slote’s central claim in his normative ethics is that ‘differences in the strength or force of our empathy make a difference to how much we care about the fate of others in various different situations,’ and that this corresponds to ‘differences of intuitive moral evaluation.’ (ECE, pp. 15-16) Slote argues (a) that there is a correspondence between the strength/force of our empathic reactions and what we take our moral obligations to be, and (b) that empathy can therefore provide a ‘plausible criterion of moral evaluation.’ (Ibid., p. 16) Slote’s empathy-based moral criterion is as follows:

Actions are morally wrong and contrary to moral obligation if, and only if, they reflect or exhibit or express an absence (or lack) of fully developed empathic concern for (or caring about) others on the part of the agent.’ (ECE, p. 31)

By analogy, we can take it that actions are morally right and not contrary to moral obligation if, and only if, they reflect or exhibit or express empathic concern for others. (Cf. Brady, 2003, p. 145) The plausibility of using empathy as a moral criterion is enhanced, Slote argues, by the frequent correlation of our empathic reactions with what we take our moral obligations to be. He argues that there are many correlations between the strength of our empathic reactions in particular circumstances, and towards particular people, and what we take to be our moral obligations towards them.
Our empathic reactions, Slote claims, vary according to several ‘modalities’. These are ‘aspects of an agent’s interaction with the world’ (ECE, p. 43) and result in empathizers feeling more or less empathic concern under certain circumstances. The modalities empathy is sensitive to are as follows.

(1) **Partiality**: our empathic reactions are strengthened or weakened according to how close the object of our empathy is to us in terms of the relationship they bear us.

(2) **Perceptual immediacy**: we feel stronger or weaker empathic concern depending on how directly we perceive the suffering of the other.

(3) **Temporal immediacy**: we feel stronger or weaker empathic concern according to when in time the suffering is occurring.

(4) **Causal immediacy**: we feel stronger or weaker empathic concern depending on whether we think another person’s suffering was caused deliberately or merely allowed.\(^8^8\)

Partiality, Slote argues, results in our feeling more empathic concern for our nearest and dearest than for strangers, and this correlates with the fact that we tend to think we ought to help those closest to us before strangers. Slote claims that we would actually disapprove of, or criticise, someone if they chose to help strangers over and above their own kith and kin. (ECE, p. 28) Perceptual immediacy, Slote argues, causes us to feel more empathic concern for those suffering right now, in front of us, than we do for those suffering on the other side of the world, and this correlates with the fact that we tend to think we have stronger obligations to help someone suffering right now.

\(^{88}\) The claim Slote makes about our empathic reactions being sensitive to certain modalities is similar to Hume’s claim that sympathetic reactions are influenced by the principles of association. In Chapter 2 (§2.2.1) we saw that, on Hume’s view, sympathy is governed by the relations of contiguity, causation and resemblance, which means that our sympathetic reactions will be partial, unless the sympathizing is done from the common point of view.
in front of us, than to help the person suffering on the other side of the world. (ECE, p. 24) Slote illustrates the temporal immediacy modality with an example in which a group of miners are trapped underground after the collapse of a mine-shaft. (ECE, p. 25) He asks us to imagine that there is only a limited amount of money to spend. Ought we to spend the money saving the currently-trapped miners, or on installing safety devices that would prevent future catastrophes? Slote claims that we would feel more empathic concern with the miners currently trapped underground than with those who might be saved if we installed safety devices that would prevent future suffering, because our empathic reactions are sensitive to when the suffering is taking place. He claims that this correlates with the fact that we tend to think we have a stronger obligation to help those currently trapped than those who might suffer in the future. Moreover, he argues, we would no doubt consider it morally reprehensible to let the interred miners die in favour of installing the safety devices that might prevent future suffering. (ECE, p. 26)

An example that showcases both the perceptual and temporal modalities is abortion. Slote claims that most people think pregnancies ought to be terminated when the mother is at risk, and that this correlates with the fact that they feel more empathic concern for her than for an unborn foetus, whose suffering is less apparent to them. (ECE, p. 17) In fact, we tend to support a woman’s right to choose, Slote argues, because we feel more empathic concern for the mother than we ever possibly could for the foetus. (Ibid.)

The causal immediacy modality accounts, Slote claims, for why we tend to think killing is worse than letting die. Our thinking in this case, he argues, correlates with the fact that empathy is more sensitive to causing harm than
allowing it. In addition to the correlation between what we take to be our obligations to help and the strength of our empathic reactions, Slote makes the further claim that our empathy can correlate with our moral intuitions about harming. We tend to think soldiers who went into villages and murdered children and civilians in cold blood during the Vietnam War were morally more culpable than those who dropped napalm on those same villages and their civilian inhabitants from the air. This is because, Slote argues, ‘the person who is willing to kill innocents in cold blood acts more unfeelingly, demonstrates a greater lack of (normal or fully developed) empathy, than the person who kills from the air.’ (ECE, p. 25) The ‘empathic reaction’ we have that explains why we think it is worse to kill than let die is what Slote calls ‘empathic aversion’ or empathic ‘flinching’. (ECE, pp. 44, 47) If an agent causes a death by killing another person—if he is a murderous soldier, say—he is ‘in causal terms more strongly connected to the death’ than if he merely allowed the death to occur, by, for example, napalming from the air. (ECE, p. 44) In other words, empathy is sensitive to the strength of the causal relation we bear to the act, with the result that we empathically flinch more from, or are empathically more averse to, causing pain to someone by harming them ourselves than we are to simply allowing them to be harmed.

Causal immediacy is supposed to operate analogously to the other modalities: just as we feel more empathic concern for those we perceive to be in pain in front of us than we do for those far away, so do we feel more empathic aversion towards causing harm than allowing it to happen. (ECE, p. 44) And just as we think we have more of an obligation to help those close to us than we do those far away, so do we think it is morally worse to kill or harm than to allow to die because we are more causally responsible when we kill than when we let die.
4.3.2 The Bias Problem

As we saw in Chapter 2, Hume considered it to be a negative consequence of sympathy that our sympathetic reactions are naturally biased. We find it easier to sympathize, for example, with those we love, and with our fellow countrymen. (See §2.4.2) Morality ought to be impartial, according to Hume, and therefore some ‘external’ device—‘external’ but still sympathy-based—like the common point of view must be invoked to ensure the elimination of any natural bias from our sympathy-based moral judgements. Slote, however, embraces the partiality of our empathic reactions. He argues not only that we do feel more empathic concern for our nearest and dearest, and that this would lead to our helping them over strangers or those far away, but also that this preferential helping is right and can be justified. It does seem safe to say that most of us consider ourselves more strongly obliged to help our nearest and dearest than strangers, when both are suffering equally in front of us. In this case, using empathy as a moral criterion—a criterion against which to judge the rightness and wrongness of actions—might not present such a problem because it would not lead to our justifying an action that we generally consider to be morally impermissible. Our empathic reactions appear to ‘track’ what we do actually take to be our moral obligations, at least in this case.

However, there are dangers with using empathy as a moral criterion. To be plausible as a basis for such a criterion, empathy—and our empathic reactions—must never end up justifying actions that we do or would consider unjustifiable. In Slote’s words, a moral criterion based on empathy must not end up ‘justifying unjustifiable distinctions’. (ECE, p. 35) If it turns out that empathy is inherently biased—if we feel more empathic concern for members of our own race, for example—then there is a danger that unjustifiable actions would come out as morally right. We do not tend to
think it is morally reprehensible to help our nearest and dearest before we help strangers. In fact, we think it is morally preferable. But we do think it is impermissible to help a member of our own racial group (an ‘ingroup’ member) before we help a member of another racial group (an ‘outgroup’ member), simply on the grounds that they are a member of a different racial group. If we feel more empathic concern for ingroup members, and less for the outgroup members, and what is right is determined by the strength or weakness of our empathic reactions, then using empathy as a moral criterion would endorse such preferential helping along racial lines, and this would be an unwelcome consequence of Slote’s position. Slote anticipates the bias problem, noting its potentially negative implications for his view:

[…] one may well wonder whether every natural distinction in or of empathy corresponds to distinctions with intuitive moral force, and perhaps the greatest of these worries concerns possible differences between the empathy we may feel for people of a different (or opposite) race, religion, or sex/gender. (ECE, p. 34)

In this section I will focus on racial bias. Slote denies that there is any significant empirical evidence of our being naturally predisposed to feel

89 There appears to be little evidence of implicit gender bias in empathy with one’s own sex. There is, however, evidence of significant gender bias in empathizing ability. A recent review study undertaken by Bhismadev Chakrabarti and Simon Baron-Cohen (2006) concluded that there are significant gender differences in empathizing ability. These differences include superior female mindreading and face-based emotion recognition abilities, as well as a tendency to be more cognizant of, and responsive to, another’s distress. (Ibid., p. 408) In addition, there is evidence that points to gender-specific differences in the mirror neuron system, thought to be an important component in empathic responding. (Cheng, et al., 2008) These findings show that the female capacity for empathy is naturally superior to its male counterpart. If women are better at empathizing, and empathizing causes empathic concern, then women will be more likely to feel empathic concern, which on Slote’s view is the fundamental moral motive. This means that, according to Slote’s moral criterion, women will be more able to act rightly and fulfil their moral obligations than men. In short, women will be morally superior to men. Slote claims in ECE that he does not consider this an unwelcome consequence of his view. He claims that ‘an ethics of care that centres around empathy needn’t feel too uncomfortable with the idea that women are now, and are likely to remain, morally superior to men; and the acceptance of that conclusion doesn’t in any way debar care ethics from functioning as a moral standard or ideal for both men and women.’ (ECE, p. 73) Whilst I do not wish to actively endorse Slote’s conclusion, I have opted to pursue the racial bias objection in this section because showing that empathy is biased along racial lines is arguably more damaging to Slote’s position. The consequence of empathy being racially biased would mean we ended up endorsing morally objectionable behaviour, if we used empathy as a moral standard. We would end up ‘justifying unjustifiable distinctions’, in Slote’s words. But the fact that women are
more empathic concern for members of our own race. (ECE, p. 35) He does quote one study which implies the existence of racial bias in children, but argues that this can potentially be explained by their upbringing, as opposed to any ‘innate predisposition’ to empathize more with ingroup members. (ECE, p. 35) However, there is a good deal of evidence demonstrating that we are disposed to feel much stronger empathic concern for, and are more willing to help, members of our own race. (Mathur, et al., 2010) There is also evidence that we can fail to feel empathic concern for members of other races, even when they are in distress. (Batson & Ahmad, 2009; Chiao & Mathur, 2010) The failure to empathize with outgroup members can lead to indifference to their suffering, and even schadenfreude, when two racial groups are in competition. (Cikara, et al., 2011; Smith, et al., 2011) And finally, the evidence suggests that the disposition to feel less empathy for outgroup members affects both low-level, associative empathy (Chiao & Mathur, 2010) and high-level, projective forms of empathy (Mathur et al., 2010), both of which Slote includes within his broad empathy concept.

Given this evidence, using empathy as a moral criterion looks highly problematic. If we feel more empathic concern for members of our own race, we will be more likely to help them before we help an outgroup member. And if acting rightly is acting according to how much empathic concern you feel for another person, and we feel more empathic concern for members of our own race, then such preferential helping will constitute right action. Right action will also include feeling indifference to, and even delighting in the suffering of, outgroup members, and failing to help members of an outgroup even when they are in distress. These are highly unwelcome

better empathizers with everyone is not as damaging to Slote’s claim that empathy can serve as a moral criterion, for it does not imply that men cannot be moral—they are, after all, capable of empathizing too—and therefore does not seem to entail such a profoundly negative consequence as implicit racial bias does.
consequences for a view that seeks to ground moral obligation and rightness in empathy. In addition, they are instances of our empathic reactions failing to correlate with what we take our moral obligations to be. The existence of such a correlation is what Slote claims lends significant plausibility to his view that empathy can ground morality.

Slote considers himself to have an alternative response to the racial bias worry, however. I have quoted the relevant passage here. Slote claims that:

> Even if there turned out to be such a predisposition in very young children, that predisposition might fade over time with the child’s increasing cognitive maturity [...] just as the early tendency simply to empathize with another person’s occurrent feelings eventually gives way, in the light of greater knowledge of the factors that affect human wellbeing, to empathic dispositions that take account of aspects of another person’s situation that transcend the person’s immediately occurrent feelings. (ECE, p. 35, Slote’s italics)

Earlier, in the discussion of Slote’s conception of empathy (§4.2), we saw that a person can empathize, on Slote’s view, with someone’s future condition, not just their occurrent feeling. He returns to the example of a terminally ill man boisterously enjoying himself, unaware of his own condition, in relation to the racial bias problem. A child, with his less-developed capacity for empathy, would be likely only to empathize with the occurrent feeling of the sick man, Slote argues, which would result in his feeling pleasure at the man’s pleasure. (ECE, p. 35) An adult empathizer, however, is likely to have a different reaction. He will be aware of the fact that the sick man will undoubtedly suffer in the future, and so ‘might well have a less positive and even a negative empathic reaction to such temporary enjoyment’ in light of this knowledge. (ECE, p. 35) Slote takes this as evidence that ‘we have little
reason to believe that people have an innate and long-term tendency’ to feel
greater empathy for ingroup members. (*ECE*, p. 35)

But what does Slote mean when he says that any ‘predisposition [toward
greater empathy with members of one’s own race] might fade over time with
the child’s increasing cognitive maturity’? There are two ways we can
interpret his statement. Firstly, we can interpret him as claiming that it is
our racial bias that is immature. It will fade over time as we become more
mature: growing up will thus make us lose any predisposition to be biased in
this way. However, this interpretation is problematic for Slote, for the
evidence presented above showing the existence of implicit racial bias in our
empathic tendencies is from *adults*, not children. These are mature
empathizers, and yet their empathy is still biased. The implicit bias has *not*
faded with increased maturity, so it would seem that racial bias is not what
is immature.

The second, potentially less damaging way to interpret Slote’s claim in the
race case is as follows. Perhaps he means that we have two empathic
reactions: an empathic reaction for individuals and an empathic reaction for
groups. Maybe it is the child’s capacity for empathy that is immature.
Children can only empathize with individuals, whereas adults can
empathize with whole groups—what fades over time is the predisposition to
only empathize with individuals. This would fit in with Slote’s claim in §4.2
that empathy’s development culminates in mature adult empathizers being
able to feel empathic concern for both individuals *and* groups. If this second
interpretation is correct, adults’ mature empathy with the future suffering of
the racial minority as a group could ‘trump’ their initial, biased empathic
reaction to the suffering of the individual member of their own race.
However, if this second interpretation is the one Slote endorses, he faces a problem, for it is not the case that our empathic concern for groups always trumps our empathic concern for individuals. Quite the opposite, in fact: empathic concern for the individual trumps empathic concern for the group in cases of what psychologist Paul Slovic (2007, p. 82) has termed ‘psychic numbing’. Slovic describes this phenomenon as a ‘fundamental deficiency in human psychology’ that arises when we become insensitive to, and fail to be motivated by, the true extent of human suffering when it is presented to us in statistics and numbers. (Ibid.) Slovic argues that psychic numbing renders us incapable of empathizing with the suffering of large numbers of people. In fact, we feel less compassion and concern the bigger the number. A devastating consequence of psychic numbing, Slovic claims, is our repeated failure as a species to act against the mass murder and genocide so endemic in the world. His argument centres on evidence demonstrating how numbers and statistics fail to convey the extent of human suffering because they cannot arouse our affective responses, with the result that we are not motivated by the suffering and plight of others. What is lacking are images and representations of suffering; because of the sheer numbers involved, it is impossible to convey the suffering of all those people such that it elicits our compassion and concern. Slovic argues that, in actual fact, we find it much easier to empathize with individual members of suffering groups than with the whole group, because the ‘identified individual victim, with a face and a name, has no peer’ when it comes to arousing our empathic concern and evoking helping behaviour. (Slovic, 2007, p. 86) Our empathy for individuals trumps our empathy with large groups because we are only able to empathize with an individual who is suffering. Where large groups of sufferers are involved, we seem unable to get our heads around the numbers, with the result that we fail to feel empathic concern and remain unmoved by their suffering. Slovic argues that perhaps the only way we can
empathize with, and be motivated by, the plight of large groups is by first of all empathizing with an individual who represents that group.

Given the fact that there are cases in which empathy for a large group fails to trump empathy for the individual, if Slote is claiming that empathy for groups always trumps empathy for individuals, his claim is false. Moreover, there are instances of what we can call ‘competing empathies’, when an empathy modality ‘trumps’ empathy for the group. Slote himself has an example in which empathy for the group gets trumped by another modality, the temporal immediacy modality. In the trapped miners case, Slote argues that his empathy-based moral criterion could deal with a situation in which we were faced with a choice between spending money on saving currently trapped miners, or on installing safety devices that would prevent future deaths. (See §4.3.1.) Slote claims that we would feel more empathic concern with the miners trapped underground right now than with the potentially much larger group who would suffer in the future if we did not install the safety devices, because our empathic reactions are sensitive to when the suffering is taking place. (*ECE*, p. 26) So in this case, the temporal immediacy modality trumps empathy for groups.

There are other examples in which empathy for the group is trumped by a modality. For example, Slote claims that we have stronger moral obligations for our nearest and dearest than we do for those far away because we feel more empathic concern for the former. This is the perceptual modality. In *ECE* (pp. 21-22) Slote argues against a view of moral obligation put forward by Peter Singer (1972), who claims that our obligations to help those suffering in far away countries are just as strong as our obligations to help those near and dear to us because distance is morally irrelevant. To Slote, this seems highly counterintuitive. He argues contra Singer that distance is
morally relevant, and that our moral intuitions appear to reflect this. The perceptual modality is invoked by Slote to explain why we think we have stronger obligations to help the child drowning in front of us than we do to help starving children in Africa. (ECE, p. 21) In this case, it is clear that there should be empathy for a whole group, and it is being trumped by empathy for an individual. Slote could simply reply that this is a case of psychic numbing: we do not feel empathic concern for the children in Africa because we are numb to their plight. But if he does say this, then he must surely accept that empathy for groups does not always trump empathy for the individual, for that is what the case of psychic numbing just presented shows.

Slote need not accept the claim that empathy for groups always trumps empathy for individuals. He could simply say that it sometimes does, and when it does is an empirical question, to be answered at some point in the future. He could therefore maintain his claim that it happens in the race case, until empirical evidence testifies to the contrary. However, if he accepts the importance of empirical evidence in this case, then Slote must also surely accept the legitimacy of the existing empirical evidence showing inherent racial bias in adults, presented earlier. Furthermore, it is plausibly false that empathy for large groups trumps empathy for individuals anyway, for there does not appear to be any evidence to show that this is the case in the literature on implicit racial bias. It does not look, therefore, as though opting for the second interpretation will enable Slote to counter the empathic bias objection in a satisfactory way, for it is not clear that empathy for groups would ever trump empathy for the individual.

In conclusion, it does not appear that Slote can solve the empathic bias problem, which means his claim that empathy is sufficient for use as a moral
criterion is false. We can therefore see that he faces aspect (2) of the motivation objection: our motives mean that empathy is not sufficient for fulfilling this function. Slote faces aspect (2) because it transpires that we are disposed to feel more empathic concern for members of our own race than members of other races. Empathy is thus insufficient for moral obligation for it would lead, for example, to endorsement of preferential helping along racial lines. And to be sufficient for moral obligation, empathy must never end up justifying actions like this.

4.3.3 Summary

In this section I have applied aspect (2) of the motivation objection to Slote’s claim that empathy is sufficient for moral obligation. I argued that using empathy to ground moral obligation is problematic because we are disposed to feel more empathy for members of our own race, so basing moral obligations on these reactions would lead to our endorsing preferential helping along racial lines. Against Slote’s claim that there is no evidence for racial bias in empathy, I presented evidence to the contrary. I then addressed his second claim that any predisposition to feel greater empathy for members of our own race would fade over time as we become more cognitively mature. I argued that his claim could be interpreted in two ways. Firstly, Slote could mean that our disposition to be biased is immature, and would fade over time. However, we saw that the evidence presented is from adults, so this interpretation will not work for Slote.

Secondly, I argued that we could interpret Slote as claiming that, when we are children, our empathy is immature insofar as we can only empathize with individuals. Adults, by contrast, can empathize with groups as well as individuals. And in the case of racial bias, Slote could argue that our empathy with groups trumps our empathy for individuals. But this
interpretation is also problematic for Slote, I argued, because it is not the case that empathy for groups always trumps empathy for individuals. Firstly, there are cases of psychic numbing in which we can fail to empathize with large groups, and can in fact only empathize with an individual who represents the large group. Secondly, there are examples of what I called competing empathies in which empathy for groups can be trumped by another modality. Slote need not accept the claim that empathy for groups always trumps empathy for individuals. He could simply argue that it sometimes does, and whether it does or does not in the race case is an empirical question. However, I argued that, if he accepts the importance of empirical evidence in this case, then he must surely also accept the importance of empirical evidence in general, including the evidence showing inherent racial bias in adults. Finally, I argued that it is plausibly false that implicit empathic bias ever trumps empathy for individuals anyway, for it is nowhere in the empirical literature on empathic bias. I concluded that Slote faces aspect (2) of the motivation objection, for his claim that empathy is sufficient for moral obligation is false.

4.4 Moral Evaluation

4.4.1 Introduction

In the previous section I applied aspect (2) of the motivation objection to Slote’s claim that empathy is sufficient for moral obligation. In this section I apply aspect (2) to Slote’s theory of moral evaluation. Slote claims that empathy is sufficient to cause us to feel sentiments of approval and disapproval when others display or fail to display empathic concern in their actions. However, I argue that, because empathy can be modulated by our beliefs and prejudices, this will affect the outcome of our empathizing, which means that empathizing is not guaranteed to cause the moral sentiments. Slote therefore faces a similar problem to Hume and Smith: empathy is not
sufficient for the moral sentiments because the outcomes of the empathizing can be affected and influenced by our motives.

4.4.2 Approval and Disapproval

Empathy mechanisms, Slote argues, cause us to have feelings of moral approval and disapproval. (MS, p. 28) They enable us to feel approval and disapproval in a similar way to Hume’s sympathy mechanism, Slote claims, insofar as they enable observers to ‘pick up’ others’ sentiments. (MS, p. 30) Slote does not explain exactly how empathy mechanisms are supposed to operate in the case of moral judgement; he refers the reader instead to Hoffman’s (2000) discussion of empathy mechanisms in the latter’s book. (MS, p. 30, footnote 5) Slote relies on what he describes as a Humean, contagion-like empathy, which he claims is essentially the capacity to be warmed (or chilled) by empathic concern (or its lack), when this is displayed (or lacking) in agential action. (MS, p. 35)

Slote borrows from Hume the basic idea that there is such a mechanism that enables the transmission of sentiments, but focuses on the transmission of the traits or motives of agents rather than the effects—the pleasure or pain—that these traits cause in others, as Hume does. (MS, p. 33) On Hume’s view, as we saw in §2.4.1, a spectator feels approval and disapproval by sympathizing with the pleasure or pain that R, the recipient, feels as a result of A’s actions.\footnote{In this chapter I will be referring to spectators as ‘S’, actors as ‘A’, and recipients as ‘R’, as previously.} According to Slote, however, it is the agent’s motives themselves that cause observers to feel approval, rather than their effects on a third party. This, Slote claims, avoids an objection—allegedly made by Smith—that Hume attempts to address in the second Enquiry. (Hume, 1777/1975, p. 213; MS, p. 32) The objection runs as follows. If it really were the case that approval and disapproval were feelings of pleasure and pain,
felt by an observer as a result of the pleasure or pain the agent’s actions or traits caused to others, we could end up approving of anything, so long as it caused pain or pleasure to those other agents. ‘Anything’ could include rocks or hurricanes, for example, as these tend to have painful effects on others. We tend, however, to think of moral approval and disapproval as being of human beings and their actions, and not of inanimate objects like rocks and hurricanes. (MS, p. 34) Keeping the contagion aspect of Hume’s theory, but also the agent-based element of Smith’s, Slote contends that we feel approval and disapproval via a contagion-like transmission of the warm or cold motives of moral agents. And the particular motive that is manifested (or not) in agents’ actions that causes approval and disapproval is empathic concern (or its lack).

Slote describes his theory of moral evaluation as causal-sentimentalist, by which he means that the feelings of approval and disapproval aroused in an observer are caused by the expression of the moral motive (or its lack) in the actions of moral agents. (MS, p. 42) Earlier (§4.2) we saw that, according to Slote, empathy mechanisms cause empathic concern in people with a normal capacity for empathy when they see or hear about someone who is suffering. This is what Slote refers to as ‘first-order’ empathy. (MS, p. 46) First-order empathy is ‘the most familiar form or instance of empathy’, and is directed at R, but felt by A for R, when R ‘needs help or is suffering.’ (MS, p. 34) Observers of the actions of others feel ‘second-order’ empathy, and this constitutes their sentiment of approbation. (Slote, 2010b, p. 136) Second-order empathy is the feeling caused in an observer by an agent’s first-order empathy for another agent, and constitutes his sentiment of approval. (Ibid.) It is ‘empathy with empathy’; that is, it is empathy ‘with’ A’s empathic concern for R. (MS, p. 35; 2010b, p. 136) First-order and second-order empathy are distinguished solely according to what causes them. (MS, p.36)
First-order empathy is caused by R’s situation, and second-order empathy is caused by A’s motive. *(Ibid.)*

A spectator’s approval, on Slote’s view, results from the following process. A acts with a motive of empathic concern towards R. S feels approval, itself an empathic reaction, to A’s empathic concern.\(^9\) If A’s action expresses empathic concern for R, S will himself have an empathic reaction aroused in him when he observes A’s action, which will be empathic warmth for A. S therefore ‘empathically reflects’ A’s empathic warmth for R, and this constitutes S’s approval of A. *(MS, p. 35)* Disapproval is to be understood by analogy with approval: if A’s action towards R exhibits a ‘basic lack of empathy, then empathic people will tend to be chilled (or at least “left cold”)’ by A’s actions. *(MS, p. 35)* S’s feeling of empathic chill, reflecting A’s lack of empathic concern for R, constitutes his disapproval of A.

Slote holds that the actions of agents are evaluated as morally good or bad ‘in terms of whether they express, exhibit or reflect empathically caring motivation, or its opposite.’ *(MS, p. 125)* Approval and disapproval are, respectively, feelings of ‘warm tenderness and coldness or chill that are part of an empathic, but to that extent also a causal, response to the motives/attitudes/feelings of agents.’ *(MS, p. 41)* They are causal responses to the *motives*, attitudes and feelings of agents, not to the effects of their actions. Slote’s claim is that if A displays empathic concern in his action towards R, then if S empathizes with A, he will feel approval. But I will now argue that, because empathic responses can be modulated by the empathizer’s beliefs and prejudices, these can affect whether or not he feels approval as a result of empathizing.

---

\(^9\) On Hume’s view, S’s approval would be felt in response to the effects of A’s action on R.
4.4.3 The Modulation of Empathy

There is evidence to show that our empathic reactions can be modulated by our knowledge and beliefs. For example, contextual appraisal involves the modulation of the empathizer’s response to the other person’s suffering by various appraisals he makes of the situation. Situational factors can include: the relationship the empathizer has with the sufferer (Singer & Lamm, 2009); the empathizer’s racial bias (Xu, et al., 2009); whether the empathizer thinks the sufferer has behaved unjustly towards her (Singer, et al., 2006), and the empathizer’s stigmatization of the sufferer (Decety, et al., 2010). One study, by neuroscientist Tania Singer and colleagues (2006), found that empathic responses are modulated by learned preferences. Male and female volunteers played an economic game in which two participants played fairly or unfairly. According to fMRI scans taken while the players watched their fellow players receiving pain, both the men and the women had stronger empathic reactions for the fair players, and the male players felt much less empathy for the unfair players than the women did. This was taken to show that the men’s empathic responses were modulated or influenced by their belief that unfair behaviour is wrong and that people who behave unfairly ought to be punished. Another study, undertaken by Decety and colleagues (2010), showed that empathizers’ empathic responses towards the pain of AIDS sufferers were modulated by their implicit biases towards, or stigmatization of, the sufferers. If, for example, the sufferers had contracted AIDS from an infected blood transfusion, empathizers felt more empathy. By contrast, less empathy was felt for those who contracted AIDS from sharing infected drugs needles. This particular study drew the conclusion that the attitudes of the empathizers towards drug abuse influenced the strength of their empathic reactions.
This evidence shows that empathic reactions can be affected by the beliefs
and attitudes of the empathizer. Slote argues that if A displays empathic
concern in his action towards R, then S will feel approval if he empathizes
with A. The spectator’s empathy can be modulated by beliefs and prejudices
about A, and these can affect the outcome of his empathy. Given Slote’s
claim that the same empathy mechanisms are involved in first- and second-
order empathizing, it seems reasonable to suggest that S’s second-order
empathy could be modulated by his beliefs and prejudices, just as A’s first-
order empathy can. They could, therefore, affect the outcome of his
empathizing—his moral sentiments. If this is the case, then it is surely
possible that whether S feels approval or disapproval of A as a result of
empathizing can be affected by his motives. For example, perhaps S is
prejudiced against A. Even if A’s action towards R does display empathic
concern, given that S’s prejudices and biases towards A can influence the
outcome of his empathizing, S is not guaranteed to approve of A’s action by
empathizing with him.

Slote could respond here by emphasising that it is only an automatic,
involuntary low-level empathy mechanism like contagion that is involved in
second-order empathy. S’s prejudices will not therefore affect whether or
not he approves or disapproves of A because this kind of empathy cannot be
modulated. There are two ways in which this response would be
unsatisfactory, however. Firstly, the evidence presented above shows that
LLE can be modulated. Secondly, insisting on only using a contagion-like
form of LLE would render Slote’s account unable to handle approval and
disapproval in a variety of circumstances. As Slote himself acknowledges, it
is possible to ‘approve or disapprove of actions and individuals that one
merely hears about and that may not actually exist or have existed (in the
way they are depicted to one).’ (MS, p. 39, footnote 23) It is hard to see how
contagion-like LLE could facilitate approval and disapproval in cases such as these, because in order to be ‘empathically warmed’ by A’s empathic concern as a result of contagion, it would have to be the case that S was present when A undertook his action. This would prevent S’s being able to approve and disapprove of people he merely hears about, of those who have existed in the past, or are fictional, and so would result in Slote’s account of approval being severely limited.92

4.4.4 Summary

In this section I applied aspect (2) of the motivation objection to Slote’s claim that empathy is sufficient to cause spectators to feel the moral sentiments of approval and disapproval. I argued that, because empathy can be modulated by our beliefs and prejudices, empathizing will not guarantee that spectators feel the moral sentiments, even if actors’ actions display empathic concern. I suggested that Slote could respond by insisting that it is a low-level contagion-like form of empathy that causes the moral sentiments, and that this is less likely to be modulated because it is automatic and involuntary. However, the evidence presented shows that, in fact, low-level

92 Several commentators consider Slote’s account of approval to be unsatisfactory for different reasons. Justin D’Arms and Angela M. Smith object that Slote holds an implausible view of the phenomenology of approval and disapproval. On Slote’s view, S’s approval has to feel warm and his disapproval has to feel cold or chilly. (MS, p. 35) D’Arms and Smith both argue (independently) that disapproval does not always seem to feel chilly. (D’Arms, 2010; Smith, 2010) Sometimes disapproval feels a lot like anger, for example, which it is harder to characterise as a cold or chilly feeling. (D’Arms, 2011, p.138) Slote has responded to this objection by arguing that anger is not a suitable characterisation of disapproval, either. (Slote, 2011) Stephen Darwall (2010) claims that Slote’s account fails to capture the intentionality of A’s motive, what A’s motive is ‘about’. Contagion-like empathy mechanisms could cause S to feel warmth by picking up A’s feeling, but S would not know either what his own feeling was caused by, or what A’s feeling was caused by. If S saw A undertake some action towards R, he may pick up A’s warm feeling, but he would not know that A’s warm feeling was one of concern about R’s well-being. So it does not look plausible, Darwall argues, to suggest that S could know what had caused him to feel empathic warmth, even if it really was A’s motive that had done so. (Darwall, 2010, p. 155; cf. D’Arms, 2011, p. 136) These are important criticisms and it remains to be seen whether Slote’s theory can handle them satisfactorily. It is at least prima facie plausible that the phenomenology objection can be answered, not least because there is a modicum of plausibility to the suggestion that disapproval feels ‘chilly’. And Slote would presumably reply to the intentionality objection that empathy always involves the maintenance of an awareness of the self as separate from the other, as we saw earlier (§4.2). I have chosen, however, to focus in this section on applying the motivation objection because I do not think it has been made against Slote’s account before.
empathy can be modulated. Moreover, if Slote were to insist on involuntary, contagion-like empathy underpinning moral evaluation, I argued, his account would be extremely limited, for it would be unable to deal with approval and disapproval in a variety of circumstances. I conclude that Slote’s claim that empathy is sufficient for moral approval and disapproval is false.

4.5 Virtue

In this section I apply aspect (3) of the motivation objection to Slote’s attempt to define virtue in terms of empathy. Aspect (3) is the circularity aspect of the motivation objection. It is the claim that, because empathy is neutral about motive, using it to define or explain certain aspects of morality in a non-circular way is problematic.

Slote, as we saw in §4.2, claims that empathic concern is the fundamental moral motive, and that acting virtuously consists in acting with this motive. He also claims that empathy mechanisms cause empathic concern. Virtue is therefore defined as empathic concern, which is a product of empathy mechanisms. However, in the previous two sections we have seen that these empathy mechanisms can be modulated by our beliefs about the object of the empathy, and by our biases, which affect the outcome of the empathy. Empathy mechanisms can, for example, be modulated by the empathizer’s bias towards, and stigmatization of, the object of the empathy. It appears, therefore, that we will only feel empathic concern for others as a result of empathizing if we are not already prejudiced or biased towards them, and if we do not stigmatize them. Not being biased or prejudiced, it is reasonable to assume, are characteristic of virtue. If this is right, then empathy mechanisms only produce empathic concern if the empathizer is already virtuous. Slote’s attempt to define virtue in terms of empathic concern is
therefore circular: if virtue is empathic concern, and feeling empathic concern requires virtue, then virtue cannot be defined as empathic concern without presupposing what it is trying to define, namely virtue.

4.6 Conclusion

In this chapter I evaluated Slote’s empathy-based moral sentimentalism, the final constitutive view I am considering in the thesis. I argued that Slote faces aspects (2) and (3) of the motivation objection in relation to three claims he makes about empathy’s moral role. These are: the claim that empathy is sufficient for moral obligation; the claim that empathy is sufficient for the moral sentiments, and the claim that empathy can be used to define virtue.

In section 4.2 I outlined Slote’s empathy concept. I argued that, although his definition of empathy is somewhat opaque, we can nevertheless interpret Slote as claiming that empathy mechanisms cause empathic concern. In section 4.3 I applied aspect (2) of the motivation objection to Slote’s normative ethical claim that empathy is sufficient for moral obligation. I argued that, because empathy is inherently biased towards members of our own race, there is a problem with basing morality on empathy, one that Slote himself anticipates: if we based moral obligation on empathy, and empathy is biased in this manner, then helping members of our own race before members of other races simply because they are members of our own race will be what we are obliged to do.

Slote denies that there is any evidence for racial bias, but I presented empirical evidence to the contrary. Slote claims that, even if empathy did turn out to be racially biased, it would not be a problem for his view because a predisposition to be biased would fade over time as empathizers become more cognitively mature. I argued that there are two ways we can interpret
Slote’s response. Firstly, as the claim that being racially biased is immature: it only happens when we are children, and the bias will fade away as we become mature adult empathizers. However, I argued that all the evidence for racial bias is from research into adult empathy, so empathic bias must still be present in adulthood. Secondly, we could interpret Slote as claiming that our empathy is immature: as children we can only empathize with individuals, but as adults we can empathize with groups as well. Our empathy for groups could therefore trump our empathy for individuals in cases where we feel more empathic concern for members of our own race.

I then argued that it is not the case that our empathic concern for groups always trumps our empathic concern for individuals, so if Slote were to endorse this second interpretation, he would face a problem. I argued that, given the phenomenon of psychic numbing, it is clear that empathy for groups does not always trump empathy for individuals. In addition, I argued that there is a competing empathies problem, whereby other empathic reactions sometimes trump empathy for groups. I argued that Slote need not accept the claim that empathy for groups always trumps empathy for individuals, for he could simply argue that it sometimes does—whether it does or not in the race case is an empirical question. However, if Slote were to accept the importance of empirical evidence here, he ought to accept it in the racial bias case too. Finally, I argued that, given the lack of evidence of empathy for groups trumping empathy for individuals, this claim is plausibly false. I concluded that Slote’s claim that empathy is sufficient to ground moral obligation is false.

In section 4.4, I applied aspect (2) of the motivation objection to Slote’s theory of moral approval and disapproval. Slote claims that empathizing is sufficient to cause us to feel sentiments of approval and disapproval when
others display or fail to display empathic concern in their actions, and that the same empathy mechanisms that cause us to feel empathic concern are also responsible for causing approval and disapproval. However, I provided evidence to show that empathy can be modulated by our beliefs and prejudices about the person doing the action, which in turn affects the outcome of our empathizing. I argued that empathizing is not therefore guaranteed to cause the moral sentiments of approval and disapproval, because the outcomes of the empathizing can be affected by our motives.

Slote could reply that the empathy process involved in approval and disapproval is a low-level, contagion-like empathy and is therefore less likely to be modulated because it is involuntary and automatic. However, I claimed that this response is not adequate for two reasons. Firstly, low-level empathy can be modulated; secondly, restricting his account of approval and disapproval to such a form of empathy would result in a theory that cannot handle approval and disapproval in a variety of circumstances. I concluded that empathy is not sufficient for the moral sentiments because it can be modulated by our beliefs and prejudices.

In section 4.5, I argued that Slote’s definition of virtue in terms of empathy is circular, and that he therefore faces aspect (3) of the motivation objection. Employing the modulation evidence, I argued that, because the outcome of our empathy can be affected by our prejudices, empathic concern cannot be used to define virtue in a non-circular way. If virtue is empathic concern, which is a product of empathy mechanisms, and empathy mechanisms are only guaranteed to produce empathic concern if the empathizer is not prejudiced, then virtue cannot be defined in terms of empathic concern without presupposing what it is trying to explain, namely virtue.
PART II

INSTRUMENTAL VIEWS
5 The Instrumental Value of Empathy

5.1 Introduction

In Part I of the thesis I considered three sentimentalist moral theories with a constitutive role for empathy. According to Hume, Smith and Slote, the moral significance of empathy lies in the fact that it is constitutive of, or foundational for, morality. I applied the motivation objection to all three theories, arguing that, because empathy is motivationally neutral, the empathizer’s motives mean that empathy processes are not reliable, and are insufficient for many of the important constitutive functions the sentimentalists claim. I also argued that empathy’s motivational neutrality means that it cannot be used to define or explain aspects of morality in a non-circular way.

Part II of the thesis focuses on empathy’s instrumental moral value: even if empathy is not considered to be foundational for or constitutive of morality, it is still possible to apportion instrumental moral value to it. This chapter applies the motivation objection to three instrumental views of empathy’s moral value. They are: firstly, the view that empathy is instrumentally valuable as a method of moral reasoning; secondly, the notion that empathy is instrumentally valuable as a means of resolving conflict, and thirdly, the claim that empathy is instrumentally valuable insofar as it produces helping behaviour. I will be applying aspect (2), the function aspect of the motivation objection, which is the claim that empathy is not sufficient for various functions because there are many ways in which our motivations can influence the outcomes of empathy processes. I also consider three further ways in which empathy connects with motive that, whilst not directly impinging on any instrumental or constitutive moral function, ought
nonetheless to be taken into account when apportioning moral value to empathy.

The chapter is structured as follows. In section 5.2 I argue that empathy is not sufficient as a method of moral reasoning. Empathy has a long history in ethics as a method of moral reasoning, manifested in, for example, the Golden Rule, versions of which can be found in many of the world’s religions. Roughly, the Golden Rule states that we ought to treat other people as we ourselves would like to be treated. Empathizing can be a way of applying the Golden Rule. If we are trying to decide what we ought to do, we can put ourselves in another person’s shoes and imagine how we would feel if we were treated in the manner in which we intend to treat the other person. The hope would then be that this process yields an outcome that tells us what we ought to do or how we ought to act. I argue that neither form of HLE is sufficient for such outcomes because the empathizer’s motives can lead to his forming the wrong conclusion as to his moral obligations. Neither will adopting the perspective of an impartial spectator mean that the empathizer comes to know what he ought to do, for HLE cannot tell him whose perspective he ought to privilege: his own (if using in-his-shoes-imagining) or the person whose perspective he is adopting (if using imagining being the other person), or the impartial spectator’s.

In section 5.3 I apply aspect (2) to the claim made recently by Simon Baron-Cohen that empathy is instrumentally valuable because it can help resolve conflicts. Baron-Cohen argues that empathizing with others can help us to understand their situations and, once we are made aware of how things look from the other’s point of view, we will be motivated to resolve our dispute with them. I argue that empathizing is not necessarily an aid to conflict resolution because it is difficult to empathize with acts we find morally
repugnant, for we sometimes form mental barriers that prevent us from empathizing. We can, however, attempt to overcome these barriers, but I argue that a separate motivation to resolve the conflict is required if we are to try doing so. Furthermore, even if we are motivated to attempt to overcome this barrier, the phenomenon of motivated inaccuracy means we can be unconsciously motivated to make inaccurate predictions of others’ mental states. This can lead to a failure to resolve conflict, if the parties in the dispute are motivated to make inaccurate predictions.

In section 5.4, I argue that empathy is not sufficient for caring. Empathizing is argued to be of instrumental value for morality insofar as it produces caring behaviour. Indeed, as we saw in Chapter 4, Slote makes a very similar claim. Empathizing is also considered to be instrumentally valuable as a means of improving the treatment of medical patients, because it can lead to more compassionate and sympathetic care. However, I argue that empathy is not sufficient for caring because our motivations can affect how, and whether, we will help or care for another when they are suffering or in need. Those exposed to high levels of suffering, like medical professionals, can become empathically over-aroused by the suffering of others, such that they can avoid caring for their patients altogether. This can be due to their failure to adequately self-regulate their emotions. It is, however, possible to self-regulate our emotions successfully, either unconsciously or deliberately, so caring can result from empathizing. But this, I argue, can be costly to the empathizer, so in order for appropriate helping to result from empathizing, the empathizer would have to be willing to bear this cost.

In section 5.5 I address some general ways in which motive might affect the instrumental value of empathy. The motivational neutrality of empathy opens it up to abuse in three ways, I argue, each of which should serve to
make us cautious about apportioning instrumental moral value to it. Firstly, empathy can be engaged in with a morally bad motive. Secondly, motive features within the act of HLE, which can affect the manner in which it is undertaken. Thirdly, our motivations can affect the imaginative processes that can occur post-empathizing.

5.2 Moral Reasoning

In this section I argue that empathy is not sufficient as a method of moral reasoning. It has been argued that high-level empathy, in the form of the Golden Rule, for example, is instrumentally valuable for morality because engaging in it can serve as a guide to moral action. The Golden Rule expresses the idea that we ought to do unto others as we would have them do unto us. (Wiggins, 1987a, p. 61) It has a long history in ethics, and examples of it can be found in most of the world’s religions, as well as moral philosophy. Secular humanism advocates the Golden Rule as the *sine qua non* of morality (MacLachlan, 2007-2011), and something akin to it forms the basis of R M Hare’s (1963) prescriptivism. It has also been linked to the Kantian Categorical Imperative, although Kant himself denied that it ought to be.

The Golden Rule supposedly captures the notion of universality, that the moral judgements we make should apply universally. That is, they ought to be applicable to everyone, regardless of his or her situation, interests, or

---

93 Hare’s universalizability thesis invokes a similar reasoning process to the Golden Rule, as Hare himself says. Hare distinguishes his view from Kantian and Golden Rule versions (1963, p. 30) but later on allies his view with the Golden Rule. (Ibid., p. 108)
94 Kant rejects the Golden Rule, but it is easy to recast the categorical imperative in terms of it, and scholars like Peter Corning (2011, p. 28) have pointed out that the categorical imperative is really just a version of the Golden Rule. Recasting the categorical imperative in terms of the Golden Rule, we would get something like: I ought not to steal because I think others ought not to steal from me. I do not want to pursue the question of whether Kant’s categorical imperative is a version of the Golden Rule here, but merely wish to note the fact that they are similar, and that this kind of thinking has a long and distinguished pedigree as a method of moral reasoning in ethics.
individual point of view. When we are conducting our moral deliberations, according to advocates of the Golden Rule, we can apply it to find out whether what we intend to do is right or wrong, or whether our past action was right or wrong. Some action will be right if we could permit it being done to us, if we were in the other person’s shoes. It will be wrong if and only if this is not the case. So the act of HLE is instrumentally valuable, according to some, because it is the means by which we can determine our moral obligations. However, I argue below, using two examples, that because the outputs of HLE can be influenced by our motivations, it is not sufficient as a method of moral reasoning. An empathizer’s motives can lead to his failing to determine his moral obligations using the Golden Rule.

5.2.1 Examples

5.2.1.1 James and Charlotte

The following is an example of an agent using the Golden Rule in an everyday instance of moral reasoning. Imagine two colleagues, James and Charlotte. James makes a sexist remark to Charlotte, which offends her. James does not think he has done anything wrong, but Charlotte is clearly upset. James wonders whether he has done something wrong, and whether he should apologise. He puts himself in Charlotte’s shoes and imagines how he would feel if Charlotte made the same remark to him. He asks himself whether, if he were her, he would be offended. He concludes that, no, he would not mind at all if someone said that to him. Charlotte is just being prissy—her anger is silly, and quite unjustified. He does not have to apologise, as he suspected, for he has done nothing wrong.

---

95 Mackie (1977, p. 90) calls this the second stage of universalization: putting oneself in the other person’s place.
This example highlights the insufficiency of this form of HLE as a method of moral reasoning. James applies the Golden Rule to see whether or not what he did was wrong, and whether he ought to apologise. This should, according to proponents of the Golden Rule, tell James that he did the wrong thing, and that he ought to apologise. For if he imagines whether he would be offended if Charlotte made the same remark to him, he should conclude that yes, he would be offended. However, James is putting himself, with his own characteristics and personality, into Charlotte’s shoes. Given his own disposition and character, he will conclude that she overreacted: if someone were to say that to him, he would not be offended. In fact, he would think it was rather funny. (Cf. Mackie, 1977, pp. 90-92) So James will not conclude that Charlotte’s offence is justified simply by applying the Golden Rule.

The Golden Rule theorist could respond that, when we apply the Golden Rule, we do not put ourselves in the other’s shoes, but rather we imagine being them. We imagine how that particular person would like to be treated. This would avoid the outcome of our empathizing being dependent on how we would feel in that person’s situation. However, this form of HLE is not sufficient to tell us what we ought to do either, as the following example demonstrates.

5.2.1.2 Anaesthetic

Consider the surgeon who is deliberating about whether he ought to perform a life-saving operation without an anaesthetic because there is none available. If the surgeon ends up not performing the operation, the patient will die. But if he operates without an anaesthetic, the patient will suffer unimaginable pain, much more than if he were allowed to die. If the surgeon imagines being the patient in order to find out what he ought to do, if he imagines how the patient will feel if he is operated on without
anaesthetic, he will conclude that the patient would suffer terribly. He may conclude that the patient will not want to have such pain inflicted upon him; he may prefer to die, given his current situation. But this is not, it is reasonable to argue, what the surgeon ought to do. The surgeon ought to operate on the patient without an anaesthetic, in order to save his life. The surgeon, if he values the feelings of the patient, will allow him to die, because he imagines that it is what the patient would have wanted. The surgeon is correctly motivated insofar as he desires to do what he imagines the patient would want. But he will not end up doing the right thing in this case, if he bases what he ought to do on HLE.

It may be argued at this point that, when we apply the Golden Rule, it is not our own perspective that we adopt, but the perspective of an impartial spectator. If this was the case, then James could adopt the perspective of an impartial spectator who would independently adjudicate between James’s perspective (qua actor) and Charlotte’s perspective (qua recipient). And the surgeon could imagine whether an impartial spectator would operate on the man without anaesthetic. However, another problem now arises: the Golden Rule cannot itself tell the imaginer whose perspective he ought to value. If James concludes from this imaginative exercise that he, James, would permit the sexist joke, but Charlotte would not, he—as the impartial spectator—must then adjudicate between these two perspectives, and decide whose is right. Similarly, even if the surgeon concludes from imagining being the patient that he ought not to perform the operation, but empathizing with the impartial spectator tells him he should operate, the Golden Rule will not tell him whose perspective is right, or which perspective he ought to value more highly. (Cf. Locke, 1981, p. 543) Faced with two competing perspectives, the imaginer must adjudicate between the two, and decide whose perspective carries more weight. An additional criterion of some kind could be appealed
to in order to enable James to apportion value to the right perspective. But this would, of course, mean that the Golden Rule was insufficient in and of itself to tell James that he ought to apologise, or the surgeon that he ought to operate, which is supposed to be its raison d’être. (Cf. Deigh, 1995, p. 762)

5.2.2 Summary

In this section I have argued that empathy is not sufficient as a method of moral reasoning because the empathizer’s motives can lead to his failing to ascertain his moral obligations. I used two examples: James and Charlotte and Anaesthetic. In the first example, I argued that in-his-shoes-imagining was insufficient to lead James to conclude that he ought to apologise for having told a sexist joke because his own motivations—his disposition and character—mean that he would not feel as Charlotte feels if in her shoes. I then argued, in the second example, that imagining being the other person is insufficient because the empathizer might conclude that he ought to do one thing, based on how the other person would want to be treated, when in reality he ought to do something else. Finally, I argued that adopting the perspective of an impartial spectator would not be sufficient to tell the empathizer what he ought to do either, because it cannot inform him which perspective he ought to adopt: the impartial spectator’s, or the other perspective.

5.3 Conflict Resolution

In this section I argue that empathy is not necessarily an aid for conflict resolution for two reasons. Firstly, it can be difficult to empathize with wrongdoers because we find what they have done morally repugnant: we can empathize with them, but doing so requires making the effort to overcome certain mental barriers, which in turn requires a motivation to resolve the conflict. Secondly, even if we are motivated to overcome these barriers, the phenomenon of motivated inaccuracy means that, in spite of our
good intentions, our motivations can cause us to make inaccurate predictions about the other person’s mental states.

Recently, Simon Baron-Cohen (2011, p. 124) has argued that empathy is ‘one of the most valuable resources in our world.’ It is a fundamental human capacity with significant real-world implications. (Ibid. 125)  

Empathy is like a universal solvent. Any problem immersed in empathy becomes soluble. It is effective as a way of anticipating and resolving interpersonal problems, whether this is a marital conflict, an international conflict, a problem at work, difficulties in a friendship, political deadlocks, a family dispute, or a problem with a neighbour. (Baron-Cohen, 2011, p. 126)

Baron-Cohen takes the Arab-Israeli conflict as a case in point. Many people on both sides of this dispute, he claims, ‘see only their own point of view and—in this sense—have lost their empathy for the other.’ (Baron-Cohen, 2011, p. 126) Baron-Cohen defines empathy as follows.

Empathy is our ability to identify what someone else is thinking or feeling, and to respond to their thoughts and feelings with an appropriate emotion. (Baron-Cohen 2011, p. 11)

On Baron-Cohen’s view, both stages—the recognition of the other person’s thoughts and feelings and the appropriate emotional response—are necessary for empathy. (Baron-Cohen, 2011, p. 11) Conflict resolution would therefore, according to Baron-Cohen, require an imaginative exercise (HLE) and an emotional response, for example sympathy with the other’s plight, both of which together comprise empathy.

See also de Waal (2010) and Rifkin (2009) for two other recent theories that Baron-Cohen claims are ‘putting empathy back on the agenda.’ (Baron-Cohen, 2011, p. 125)
5.3.1 Barriers to Empathy

Adam Morton (2011) argues that we find it difficult to HLE with people who commit evil acts because our sense of decency constitutes a barrier in our minds that sometimes stops us being able to imagine how they could have acted as they did. He draws a distinction between imagining how and imagining why. Take the recent case of Josef Fritzl, the Austrian septuagenarian who kept his daughter locked up in a cellar for 24 years, repeatedly raping her, fathering seven children by her, one of which he admitted to having killed by neglecting it. A court in Austria sentenced Fritzl to life imprisonment, finding him guilty of rape, incest, murder and enslavement. We cannot, claims Morton, understand how someone like Fritzl can have done what he did because our sense of moral decency prevents us from doing so. We can understand why he did it, to some extent at least, because ‘we can describe the motives and we can often even imagine some of what it might be like to do the acts.’ (Morton, 2011, p. 321) But we cannot, according to Morton, easily empathize with Fritzl because doing so requires that we overcome the barrier that prevents us from understanding how he could have done what he did. This barrier, however, is deliberately put in place by us because we do not want to believe of ourselves that we can empathize with, and thereby perhaps even come to understand, how people like Fritzl can act as they do.

Applying this to the Arab-Israeli conflict, it is plausible to argue that a Palestinian whose son has been killed by Israeli soldiers, or an Israeli whose son has been killed by a Palestinian rocket, would be unable to empathize with each other because their individual senses of decency constitute barriers in their minds, which result in their being unable to imagine how the other’s son could have done what he did. Granted, they may be able to understand why each side did what they did—their motives—but a more deep-seated
sense of moral outrage could prevent them ever being able to imagine *how* the other side could have acted that way. After all, the differences between the Israelis and Palestinians are profound and seemingly intractable, stemming in large part from fundamentally differently religious beliefs. Perhaps both parties will be unable to empathize with each other because they will be unable to understand how the other side could have done what they did.

Baron-Cohen could respond that it is possible to overcome these barriers. Indeed, according to Morton (2011, pp. 321-325), the barriers to empathizing can be overcome using various mental techniques, such that we can end up empathizing with those we consider to be morally abhorrent, should we so wish. Seeking to overcome the barriers to empathy, however, will require making an effort to do so. It seems reasonable to infer from this that people must be motivated to resolve their conflicts before they engage in the empathy, if they are interested in overcoming the barriers that might prevent their empathizing. The implication of this is that empathy is not necessarily an aid to conflict resolution, and what is actually at work is an antecedent desire to resolve the conflict.

Perhaps Baron-Cohen could accept that we require a motivation to resolve the conflict that is prior to the empathizing itself, for it would not detract from his claim that ‘any problem immersed in empathy becomes solvable.’ (Baron-Cohen, 2011, p. 126) However, he faces a further problem, even if the parties do have a motivation to resolve the conflict. For even if they are so motivated, and therefore do overcome the barriers that might prevent empathizing, the phenomenon of motivated inaccuracy could cause them to make inaccurate predictions about each other’s mental states, which could lead to a failure to resolve the conflict.
5.3.2 Motivated Inaccuracy

Even if they are motivated to empathize and overcome the barriers that would prevent their empathizing, parties in a conflict could be unconsciously motivated to come to the wrong conclusion about each other’s thoughts and feelings. Empathizers can have reason to make inaccurate predictions—what William Ickes (1997) refers to as ‘inferences’—about each other’s thoughts and feelings. Identified by Ickes, the phenomenon of motivated inaccuracy occurs when an empathizer has an underlying reason to make an incorrect prediction about another person’s thoughts and feelings because correctly predicting will have a negative impact on the relationship he or she has with another person. (Ickes & Simpson, 1997)

In threatening situations that have the potential to destabilize and destroy the relationship [...] perceivers who have much to lose in potentially destabilizing situations can avoid inferring the damaging thoughts and feelings of their relationship partners. (Ickes & Simpson, 1997, p. 243)

So, even if they are empathizing because they are seeking to resolve a conflict, the parties involved may be unaware of certain factors that might lead them to make inaccurate predictions about each other’s mental states. An extreme example of motivated inaccuracy concerns domestic violence. There is evidence that men who are prone to spousal abuse have a tendency to incorrectly predict that their wives are ‘harbouring critical and rejecting thoughts and feelings’ about them. (Ickes, 2009, p. 61) By ignoring how their wives actually feel, even in the face of direct contradiction from the wives themselves, the husbands can continue to use these biases as an excuse to perpetuate their violent behaviour, in order to maintain control within the relationship. (Ibid.) Alternatively, a person might be motivated to inaccurately predict that another person is feeling fine, when in fact he is not, for example incorrectly assuming that one’s child is feeling happy about
going to school on his first day. Correctly attributing dread and anxiety would cause you to scoop your child up and take him home. We might also, inaccurately, attribute feelings of benevolence and kindness to another person, because we want to see it that way; seeing them as they really are will mean coming to terms with the fact that they have been deceiving us for all these years.

Applying this to the conflict resolution case, it seems reasonable to conclude that, even when the parties wish to overcome the barriers to empathizing because they desire to resolve their conflict, they could still fail to resolve the conflict if they are unconsciously motivated to make inaccurate predictions about the other party’s mental states. Perhaps the father of a Palestinian boy killed in an Israeli airstrike would consider himself highly motivated to understand how the Israeli father whose son was blown up in a rocket attack feels. But he may be unconsciously motivated to infer that the Israeli father does not want to resolve the conflict with him because he (the Israeli) despises him—in spite of what the Israeli father himself says to the contrary. For we may speculate that it may be safer for the Palestinian to retain this belief than to go out on a limb and believe something that has previously been anathema to him.

5.3.3 Summary

In this section I applied aspect (2) of the motivation objection to Baron-Cohen’s claim that empathy is instrumentally valuable as an aid to conflict resolution. I argued that empathizing will not guarantee that we are motivated to resolve conflicts for two reasons. Firstly, it is difficult to empathize with acts we find morally repugnant, because we form mental barriers that prevent us from empathizing. When this occurs, a separate motivation to resolve the conflict is required if we are to attempt to
overcome these barriers. Secondly, even if we are motivated to overcome the barriers, we can be unconsciously motivated to make inaccurate predictions of others’ mental states, which could lead to a failure to resolve conflicts if the parties in the dispute are so motivated.

5.4 Empathy and Caring

5.4.1 Introduction

Another way in which empathy is seen to be of instrumental value for morality is in its ability to produce helping or caring behaviour, directed at others when they are suffering or in need. One of Slote’s central claims, as we saw in Chapter 4 (§4.2) was that if we perceive or imagine the suffering of others by either LLE or HLE, we will care about their well-being and be motivated to help them if they are in need or suffering. Empathy is often encouraged as a means of improving the care and treatment of medical patients, because understanding a patient’s situation—what it is like for them—and feeling compassion for their plight are thought to be necessary for the provision of effective and sympathetic care. (Rousseau, 2008)

However, I argue in this section that empathy is not sufficient for caring because our motivations can affect how, and whether, we will help or care for another when they are suffering or in need. Those exposed to high levels of suffering, such as medical professionals, can end up feeling over-aroused by the suffering of others if they empathize with them, to the extent that they fail to care for them altogether, if they are not able to adequately self-regulate their own emotions.

Advocates of empathy’s instrumental value for caring and helping could respond by pointing out that it is possible to self-regulate our empathic responses to the suffering of others so that we do not become over-aroused by their suffering. Empathizers can self-regulate their emotions
unconsciously, if they are experienced medical professionals, or deliberately, by emotion suppression. However, I argue that choosing to self-regulate can come at a significant cost to the empathizer, which means that, in order for appropriate helping to result from empathizing, the empathizer has to be willing to bear this cost, which will require an independent desire to do so.

5.4.2 Compassion Fatigue

Jean Decety (2011, p. 103) argues that empathizing with patients can be a ‘double-edged sword, facilitating care and compassion but at the same time leaving the physician vulnerable.’ There is evidence that consistent exposure to the suffering of others causes high levels of vicarious traumatisation, otherwise known as ‘burnout’ or compassion fatigue, in doctors and other care practitioners. (Palm, et al., 2004) Burnout can result from HLE or LLE, or from a combination of these: when we immediately recognize (by undergoing some kind of LLE process) or imagine (by HLE) another person’s pain or suffering, there is evidence that we can experience similar physiological reactions in our own bodies (Decety, 2011; Singer & Lamm, 2009; Preston & de Waal, 2002)\(^97\) which leads to empathic over-arousal and results in burnout.

Burnout leads to deficiencies in care provision because the carer is emotionally exhausted, unable to allocate precious emotional resources to his patients. Over-exposure to another’s suffering can also lead to personal distress rather than empathic concern, which can lead to care-givers avoiding caring for their patients altogether. Evidence from psychology (Hoffman, 2000) shows that HLE can be elicited by the observation of the suffering of others, and it is well documented that such perspective-taking

\(^97\) See Decety (2011) for a comprehensive review of the possible neurological mechanisms underlying our responses to perceiving or imagining another’s pain. See also Iacoboni (2011) and Goldman (2011).
processes can lead to personal distress or empathic concern. (Batson, 1991; Eisenberg & Eggum, 2009, p. 72; Hoffman, 2000)

5.4.3 Top-Down and Bottom-up Processes

What we end up feeling as a result of empathizing is determined by a complex interplay between so-called ‘top-down’ and ‘bottom-up’ processes, as well as a variety of personal and situational factors. This interplay between top-down and bottom-up factors, and personal and situational factors, in turn affects how we behave in response to the other person’s pain and suffering. (Singer & Lamm, 2009; Decety, 2011; Eisenberg & Eggum, 2009)

5.4.3.1 Bottom-Up Processes

Bottom-up processes involve the communication of affect, which consists of direct matching between the empathizer’s perception of the target’s pain and the activation of the empathizer’s own pain networks. (Decety, 2011, p. 93; Singer & Lamm, 2009, p. 8) Bottom-up processes include: LLE processes like emotional contagion or facial mimicry; what psychologists often refer to as ‘affective’ empathy (Eisenberg & Eggum, 2009, p. 73), and mirroring.

5.4.3.2 Top-Down processes

Top-down processes can affect bottom-up responses to suffering by either dampening them down or intensifying them. (Singer & Lamm, 2009, p. 11) In top-down processing, ‘the perceiver’s motivation, intentions, and self-regulation influence the extent of an empathic experience, and the likelihood of prosocial behaviour.’ (Decety, 2011, p. 93) Top-down processes include emotion self-regulation and contextual appraisal. (Singer & Lamm, 2009, p. 8)

---

98 Top-down processes play an important role in Chapter 6. (See §6.4.1.)
5.4.3.2.1 Contextual Appraisal

We saw evidence of contextual appraisal in Chapters 3 and 4 (§3.3.2 and §4.4.3). In brief, contextual appraisal involves the modulation of the empathizer’s response to the other person’s suffering by various appraisals he makes of the situation. Such situational factors include the relationship the empathizer has with the sufferer (Singer & Lamm, 2009), the empathizer’s racial bias (Xu, et al., 2009), whether the empathizer thinks the sufferer has behaved unjustly towards them (Singer, et al., 2006), and the empathizer’s stigmatization of the sufferer. (Decety, et al., 2010) In this latter case, research has shown that bottom-up responses to the pain of AIDS sufferers is modulated by beliefs about how they came to have the disease. For example, we tend to have a weaker empathic response if we think a sufferer caught the disease by sharing needles than if we think he caught it from a botched blood transfusion. (Decety, et al., 2010, p. 994)

5.4.3.2.2 Emotion Self-Regulation

Self-regulation is the ability to regulate and control our own emotions. (Decety, 2011, p. 102) It is thought to be essential to regulate our responses to others’ suffering so we do not end up feeling personal distress when we are exposed to it. (Eisenberg & Eggum, 2009, p. 73; Decety, 2011, p. 102) Eisenberg and Eggum (2009) argue that whether someone is likely to feel either personal distress or sympathy as a result of perceiving or imagining another’s pain varies according to how good he or she is at self-regulating his or her emotions, and how prone to negative emotions he or she is. (Eisenberg & Eggum, 2009, p. 75)

99 Emotion self-regulation bears an interesting similarity to Smith’s account of virtue, explained in §3.5.
Combined with good effortful control—being able to focus and shift our attention—good self-regulation is thought to lead to an empathizer’s feeling compassion and sympathy, rather than personal distress, regardless of how prone to negative emotions he is, or how intense the sufferer’s pain is. (Eisenberg & Eggum, 2009, p. 75) Conversely, as is the case for the Junior Doctor in the following example, if the empathizer is not a good self-regulator, and is prone to feeling intense negative emotions, he is more likely to feel personal distress than empathic concern or sympathy, which will mean he is more likely to avoid exposure to the suffering of the other person rather than helping him. (Ibid.)

5.4.3.3 Junior Doctor (1)

Tim is a junior doctor doing a rotation in Accident and Emergency. One of Tim’s patients, Al, is a burns victim. He is in agony, screaming with pain almost constantly. Tim knows he has to turn Al over in bed, or Al will get bed sores. But even the slightest movement causes Al indescribable pain. Being near to Al causes Tim intense emotional distress. He cannot face seeing him on his rounds, even though he knows he must do so. He cannot face turning Tim, which he must do in order to prevent bed sores from developing. As a consequence, Tim avoids Al when he does his rounds, and his neglect causes Al to develop the sores.

The ‘right’ response for Tim to have to Al’s suffering is to ‘resonate and understand [him] without becoming emotionally overinvolved in a way that can preclude effective medical management.’ (Decety, 2011, p. 103) Tim fails to have the right response in this case. His empathy is, however, correctly motivated: he wants to be a good doctor, to do a good job, to provide his patients with the best care he can. He is fully aware of how important compassionate caring is, and how empathizing can lead to his understanding
and appreciating what it is like for his patients. But empathizing leads to avoidance behaviour in this case, rather than the helping behaviour that would alleviate Al’s suffering.

Tim could self-regulate by suppressing his own emotional responses. Empathizers are able to dampen down (or ‘down-regulate’) their bottom-up responses using top-down self-regulation. This can be unconscious, occurring in virtue of the empathizer’s medical knowledge, for example. (Cheng, et al., 2007) Preliminary neuroscientific studies have shown that medical expertise dampens down bottom-up empathic responses, which is thought to explain how experienced medical professionals are able to administer effective care in the face of intense suffering, and to avoid burnout. (Decety 2011, p. 104) But as an inexperienced doctor, it is unlikely that Tim will be able to unconsciously self-regulate using extensive medical expertise. Self-regulation can also be deliberate, however, occurring in the form of emotion suppression, for example, which involves deliberately down-regulating one’s affective responses. (Butler, et al., 2003) If Tim made an effort to suppress his negative emotional responses, he might be less adversely affected by Al’s suffering, and would therefore be able to care for him as required.

5.4.4 The Cost of Self-Regulation

Those who are constantly exposed to high levels of suffering, like nurses, doctors and other caring professionals, can—and indeed must—self-regulate their empathic responses, not only in order to prevent over-arousal that might lead to burnout, but also to facilitate adequate care-giving. (Decety, 2011, p. 103) Doing so can result in empathizing producing caring behaviour. Eisenberg and Fabes (1992) have shown that over-exposure to another person’s suffering can lead to personal distress, whereas ‘optimum’ exposure
tends to lead to empathic concern, sympathy and compassion. Adequate self-regulation can lead to optimum rather than over-exposure to the other’s suffering, which will be more likely to result in caring rather than avoidance behaviour.

However, evidence shows that conscious emotion self-regulation tends to be costly for the empathizer. As Decety (2011, p. 103) argues, ‘active (conscious) regulation of negative emotions has physiological and sociopsychological costs.’ It can lead, for example, to physiological consequences such as raised blood pressure, and also social consequences such as decreased rapport with, and mutual feelings of stress and anxiety between, the empathizer and the target. (Butler, et al., 2003) If it is possible for Junior Doctor to self-regulate his emotions by consciously suppressing his negative empathic affect, he would be more likely to be optimally aroused, rather than over-aroused, by the suffering of his patients. He will then be able to care for them appropriately. But deliberate self-regulation will require a willingness on his part to bear the cost of the emotion suppression. It is reasonable to suggest that he must care about caring sufficiently to be willing to empathize and endure emotion suppression’s costly effects.

5.4.5 Summary

In this section I have argued that empathy is not sufficient to guarantee that the empathizer responds in the right way to the suffering of another person. In fact, HLE requires regulation, as well as certain facts about the empathizer and the situation to obtain, if it is to produce the right response to the other’s suffering. However, self-regulation can be costly to the empathizer: some helping will be at a price. Maintaining our desire to help others in the face of
our own pain and suffering will not be easy, regardless of whether we are correctly motivated to do so or not.\textsuperscript{100}

5.5 The Abuse of Empathy

In this final section I argue that there are three additional ways in which motive interacts with empathy, in particular high-level empathy, that should be considered whenever its instrumental (and also constitutive) moral value is being advocated, for they highlight important ways in which empathy can be abused. Firstly, motive can be an input of the empathy process: a person can empathize for morally bad reasons. Secondly, the manner in which we empathize is significant: we can empathize in order to delight in another’s suffering. And thirdly, there are various ways we can imagine after empathizing that may be morally questionable. I argue that all these aspects must be taken into account when apportioning moral value to empathy.

5.5.1 Examples

5.5.1.1 Torture

A person’s motive for engaging in HLE can be bad. The familiar example of the ‘empathic’ torturer shows how HLE can be engaged in with a bad motive. In George Orwell’s \textit{1984} (Orwell, 1949/2004), O’Brien is about to torture Winston Smith in Room 101. Room 101 is where citizens are subjected to torture by being forced to live out their worst nightmare. In Winston Smith’s case, rats represent his worst nightmare. O’Brien intends to break Winston, to make him love Big Brother. Of course, given the omniscience of Big Brother, O’Brien knows that rats are Smith’s biggest fear. O’Brien could HLE to find out the most effective means of administering the rats. He puts himself in Winston’s shoes and imagines, if he, O’Brien, feared

\textsuperscript{100} The importance of self-regulation will feature further in Chapter 6, where I discuss the role of empathy in the virtue of compassion.
rats above all things, what the worst possible way of being exposed to them would be.

Alternatively, O’Brien could imagine being Winston, about whom he knows a great deal.\textsuperscript{101} Winston himself would most fear having the rats gnawing at his face, scratching at his eyes. Whichever means O’Brien chooses, the point is that he could engage in the HLE for this patently immoral reason, and come to the conclusion that the wire cage with the mask attached is most likely to break poor Winston Smith.

It was an oblong wire cage with a handle on top for carrying it by. Fixed to the front of it was something that looked like a fencing mask, with the concave side outwards. Although it was three or four metres away from him, he could see that the cage was divided lengthways into two compartments, and that there was some kind of creature in each. They were rats. (Orwell, 1949/2004, p. 296)

A more complex contemporary variation on this example would be the ‘empathic’ interrogator at Guantanamo Bay who deliberately imagines being one of his detainees, with the motive of finding out how best to make him talk. The interrogator’s motive is to find out the best and most effective means of extracting information from the detainee, which might have been a good thing to try to do, as it would save many innocent lives. But still, we insist that torture is wrong, whatever good might come of it, and in this sense engaging in the HLE act is morally wrong as it facilitates a wrongful act. Perhaps the interrogator means well: he wants to save lives. Or perhaps he is just doing his job—orders are orders, after all.

\textsuperscript{101} See Goldie (2000, p. 195) for characterization as one of the necessary conditions for empathy.
5.5.1.2 Dentist

According to Kendall Walton (1997, pp. 44-45), our imaginings, ‘often reflect actual attitudes, desires, values, prejudices and so forth, and are thus subject to esteem and repudiation.’ Walton’s well-known example of spontaneous imagining involves imagining kittens being tortured.

[...] fantasizing about torturing kittens may, depending on the circumstances, indicate a cruel nature as surely as actually doing so would [...] If I read a story about kittens being tortured, the mere fact that I imagine this probably does not, in this case, reflect badly on my moral character. If I should find myself imagining it with a sense of glee, however, I may have reason to worry. The glee is real. But my experience certainly does not have to be described as actually taking pleasure in the suffering of kittens, in order to signal a cruel streak in my character. It is the manner in which I imagine that is significant, in this case. But in other cases, what I imagine [...] may reflect on my character. Suppose, for instance, I find myself imagining kittens being tortured, on seeing the word “kitten” on a spelling test, or simply out of the blue. Our real selves make themselves felt in what we imagine, as well as in what we feel and the manner in which we imagine what we do. (Walton, 1997, pp. 44-45, Walton’s italics)

Walton’s example shows several ways in which imagining can reveal our motivations to us. The following aspects of imagining can be revelatory of our motivations: firstly, the content of the imagining—what we actually find ourselves imagining; secondly, the way in which we imagine—gleefully, for example; finally, what triggers our imagining—the word in the spelling test.

Empathizing can enable us to delight in the pain of others. Consider the vicious dentist who takes pleasure in the pain of his patients, even though it is no part of his intention to inflict pain on them. He likes to imagine being his patients, and the pain they are undergoing at his behest. Although he is engaging in HLE with a bad motive, this case is different from the Torture
example because the empathizing is not being engaged in to inflict pain on the patients. It is, rather, the manner in which the dentist empathizes that is relevant here. He does it lingeringly, and with relish. He delights in imagining being his patients, suffering the pain he is (not intentionally) causing them.

5.5.1.3 Lolita

There are various ways things can go after we HLE, either deliberately or spontaneously: our imagination can run away with us;\(^{102}\) we can deliberately carry on our imagining, or we can intervene in it by stopping it altogether (although we might not always succeed). Our imagining can be subject to our control, as Walton (1990, p. 14) observes:

> [...] the imaginer may have the option of intervening deliberately in her imaginative experience even if she chooses not to exercise this option.

This gives us the option of trying to ‘control’ the imaginative process if we wish. The fact that it can be subject to our control means that we can also choose not to intervene when we either set out on an imaginative project, or find ourselves doing so. If setting out on the imaginative project is morally suspect, as empathizing with Humbert Humbert from Nabakov’s *Lolita* arguably is, then indulging it, allowing it to run on in our minds, is likely to be questionable, too. Wollheim’s (1984, pp. 81-82) example of sexual arousal highlights the problem.

I shall use the familiar phenomenon of the erotic daydream...Let us suppose that I centrally imagine myself with a strange figure or a close friend. As I do so, I centrally imagine myself becoming excited over what occurs between us [...] And as I centrally imagine myself becoming excited, so I become excited.

\(^{102}\) Cf. Walton’s (1997, p. 39) spelunking example.
If we were to HLE with Humbert, imagining him perpetrating one of his depraved paedophilic acts, we might find ourselves becoming aroused by the imagining. Of course, if we realise once we have started imagining that we are getting into dangerous waters and stop, this can be a good thing: we may be shocked and appalled by what we find ourselves imagining. But we might not, and this possibility should make us cautious about applying instrumental value to HLE.

5.5.2 Summary

In this section I argued that there are three additional ways in which motive interacts with empathy, in particular high-level empathy, that should be considered whenever its instrumental (and also constitutive) moral value is being advocated, for they highlight important ways in which empathy can be abused. Firstly, motive can be an input of the empathy process: a person can empathize for morally bad reasons. Secondly, the manner in which we empathize is significant: we can empathize in order to delight in another’s suffering. And thirdly, there are various ways we can imagine after empathizing that may be morally questionable. All these aspects require consideration when apportioning moral value to empathy.

5.6 Conclusion

In this chapter I utilised examples and current empirical evidence, where these were relevant, to apply the motivation objection: I argued that consideration of motive is vital when making claims for empathy’s instrumental moral value. I applied aspect (2), the function aspect, of the motivation objection to three instrumental views of empathy’s moral value. These were: firstly, the view that empathy is instrumentally valuable as a method of moral reasoning; secondly, the notion that empathy is instrumentally valuable as a means of resolving conflict, and thirdly, the claim that empathy is instrumentally valuable insofar as it produces helping
behaviour. I also argued that the empathizer’s motivations can lead to empathy being abused.

In section 5.2 I argued that empathy is not sufficient as a method of moral reasoning, for example as a way of applying the Golden Rule. I argued that neither form of HLE is sufficient for this function because the empathizer’s motives can lead to his forming the wrong conclusion as to his moral obligations. I also argued that adopting the perspective of an impartial spectator will not be sufficient for application of the Golden Rule either, for HLE cannot tell the empathizer whose perspective he ought to privilege: his own (if using in-his-shoes-imagining) or the person whose perspective he is adopting (if using imagining being the other person), or the impartial spectator’s.

In section 5.3 I applied aspect (2) to Simon Baron-Cohen’s claim that empathy is of instrumental value as a means of resolving conflicts. I argued that empathizing is not necessarily an aid to conflict resolution because we find it difficult to empathize with acts we find morally repugnant, as a result of having formed mental barriers that prevent us from empathizing with perpetrators of such acts. If the parties in a conflict consider the other to have undertaken a morally repugnant act, HLE may not be possible. Attempts can be made to overcome these barriers to empathy, but I argued that a separate motivation to resolve the conflict is required if we are to attempt this. In addition, even if the empathizer is motivated to resolve the conflict, and therefore to overcome the barrier, the phenomenon of motivated inaccuracy means he can be unconsciously motivated to make inaccurate predictions about the mental states of the other party in the dispute, which may lead to a failure to see things from his point of view.
In section 5.4, I argued that empathy is not sufficient for caring because our motivations can affect how, and whether, we will help or care for another when they are suffering or in need. Repeated and prolonged exposure to high levels of suffering can lead to empathizers becoming over-aroused to the extent that they can experience burnout or compassion fatigue, and can seek to avoid caring for their patients altogether. This can be due to their failure to adequately self-regulate their emotions. I argued that it is possible to successfully self-regulate our empathic responses, either unconsciously or deliberately, such that caring does result from empathizing with others in severe pain. But this, I argue, can be very costly to the empathizer, which means that, in order for appropriate helping to result, the empathizer must be willing to bear this cost.

Finally, in section 5.5 I addressed three general ways in which motive might affect the instrumental moral value of empathy. The motivational neutrality of empathy opens it up to abuse in three ways, I argued, each of which should serve to make us cautious about apportioning instrumental moral value to it. Firstly, empathy can be engaged in with a morally bad motive, as was the case in the example of the so-called ‘empathic’ torturer. Secondly, the empathizer’s motives can feature within the act of HLE, affecting the way in which it is undertaken. Lastly, I argued that our motivations can affect the imaginative processes that can ensue after we have empathized: we can allow our imaginations to continue down an immoral path. A process that can be abused in such a fashion is not one that we should readily consider morally valuable.

Taken together, Chapters 2, 3, 4 and 5 constitute significant evidence against both constitutive and instrumental views of empathy’s moral value. However, they should not be taken to imply that empathy can play no
positive role in morality. On the contrary: in the next chapter, I defend a particular interpretation of an instrumental view of empathy’s moral value. I argue that, when properly integrated with virtue, in particular the virtue of compassion, empathy can play an important, instrumentally valuable, role in our moral lives. Chapter 6 will also serve to justify the circularity aspect of the motivation objection that I made against the sentimentalists; I argue that virtue must precede empathy, in order for empathizing to be for the good.
6 Empathy and Virtue

6.1 Introduction

In this chapter I defend an instrumental view of empathy’s role in ethics. In Chapters 2, 3, 4 and 5, I argued that empathy’s advocates, including both those who claim a constitutive relation between empathy and ethics, and those for whom empathy has instrumental moral value, face the motivation objection because they fail to take into account the fact that empathy can be affected and influenced by our motivations. Empathy is, I argued, motivationally neutral, which has implications for the sentimentalists’ constitutive views of empathy, and the instrumental views discussed in Chapter 5, because it renders empathy insufficient for many of its purported functions.

I argued that one important function for which empathy is insufficient is compassion: because the outputs of empathy mechanisms, of which compassion is one example, can be affected by the motivations of the empathizer, empathy cannot be sufficient for compassion. In Chapter 5 I argued that empathy is not sufficient for caring, which is of particular relevance to the current chapter. I also argued that the motivational neutrality of empathy renders it open to abuse by those with bad motives. In addition, I argued that empathy’s motivational neutrality means that the sentimentalists face the reliability and circularity aspects of the motivation objection. The former is the claim that empathizing does not always lead to the right results because it is influenced and affected by our motivations. The latter is the claim that, because empathy is neutral about motive, using it to define or explain certain aspects of morality, for example virtue, in a non-circular way is problematic.
Tying these strands together, I conclude from Chapters 2, 3, 4 and 5 that, because of its motivational neutrality, only when we are already virtuous can our capacity for empathy be depended upon to be reliably for the good. Although this appears to be a negative conclusion about empathy and the role it can play in ethics, quite the reverse is true: the current chapter defends a particular interpretation of an instrumental view of empathy’s moral value according to which empathy can be a good thing—in the sense of being reliably for the good—when it is integrated in the right way with virtue. One such virtue it can be integrated with is compassion. I argue that, in the hands of the virtuous person, empathy can play a positive role in morality by enhancing this virtue.

The chapter is structured as follows. In section 6.2, I provide an outline of virtue and the virtuous person. I argue that the virtuous person will possess the virtues, central to which is practical wisdom, and also a wide range of skills, one of which is the capacity for empathy. In section 6.3 I introduce Lawrence Blum’s notion of compassion, and argue that it is superior to the notions put forward by Hume, Smith and Slote because it requires a concern for the good of the other person as a necessary condition for this virtue. I argue that the sentimentalist notions of compassion, according to which compassion incorporates, but is not preceded by, a concern for the other’s good, are inadequate because empathizing is not sufficient for concern, and therefore cannot be sufficient for compassion. I also argue that Blum’s notion of compassion lacks a role for LLE and imagining being the other person. In section 6.4, I argue that all three forms of empathy (in-his-shoes-imagining, imagining being the other person, and LLE) can enhance

103 Compassion is just one virtue where empathy plays an important role. Martin Hoffman (1990; 2000; 2011) has argued persuasively that empathy can play an important role in justice, for example. Hoffman claims that empathy can contribute positively to justice by motivating individuals to change laws they think are unfair or unjust, and by contributing to the making, changing and application of laws in the courts.
compassion. Using examples, I argue that the capacity for empathy can be integrated with virtue to enhance compassion in a variety of ways.

6.2 Virtue and the Virtuous Person

In attempting to sketch an outline of the virtuous person, and what virtue consists in, it will be helpful to take as a starting point an influential definition of virtue put forward by Linda Zagzebski:

A virtue, then, can be defined as a deep and enduring acquired excellence of a person, involving a characteristic motivation to produce a certain desired end and reliable success in bringing about that end. (Zagzebski, 1996, p. 137)

6.2.1 Motives

The virtuous person must have the right motive, one that is appropriate to the virtue. This is included in the second element of Zagzebski’s definition. Virtue involves having a characteristic motivation to produce a certain desired end and reliable success in bringing about that end. The desired end is usually taken by virtue theorists to be eudaimonia or human flourishing. We value and approve of the possession and exercise of the virtues to the extent that they contribute to this flourishing. I will assume what I take to be relatively uncontroversial: that the desired end of the virtues is, in effect, to enhance overall human well-being. (Cf. Foot, 1997) The virtues aim, as Zagzebski puts it, at ‘making the world a certain kind of place—a better place, we might say, or the kind of place good people want it to be.’ (Zagzebski, 1996, p. 137)

A virtue is deep and enduring because it is a stable and reliable character trait, acquired and shaped over time through upbringing, moral education and habituation. (Zagzebski, 1996, p. 106; Annas, 2006, p. 517) The virtuous
person must exercise the virtues reliably and habitually; it is not sufficient for being endowed with the virtue of kindness, say, that I merely act kindly on one occasion, or that I act kindly inadvertently or accidentally. Rather, possession of this virtue requires that I normally act kindly—in situations where kind behaviour is appropriate. (Wiggins, 1987b, p. 243) Similarly, the fair person will have a characteristic motivation to ‘see others treated equitably, and this leads him to want to produce a state of affairs in which the relations among people have this characteristic.’ (Zagzebski, 1996, p. 131)

I use ‘motive’ here in the broad sense outlined in Chapter 1 (§1.1). There, motive referred to the reason why the person acted. The generous person’s motive will be generosity. That is, he will have acted out of generosity, because he has a generous character, because he is a generous person. (Cf. Statman, 1997, p. 12) The fair person’s motive will be a sense of fairness. The virtuous person’s motive need not have been his occurrent thought that he ought to act in such and such a way because it is the kind or fair or generous thing to do. Rather, his motive will stem from his virtuous character, which will, as Hursthouse (1999, p. 160) puts it, go ‘all the way down’.

6.2.2 Dispositions

Whilst we often talk as if a virtue, such as compassion, is a single psychological disposition—to have the right thoughts and feelings, and to do the right thing—it is, in fact, psychologically highly complex. For example, virtuous character traits like compassion do not refer to just one disposition. To attribute a trait like compassion, generosity or kindness to a person is rather to refer ‘not to a single disposition, but to a complex network of dispositions which interlock and dynamically interrelate in ways that enable the agent both to recognise and to respond to a situation as embedded in a
complex narrative which includes the agent, and his thoughts, feelings, and actions.’ (Goldie, 2000, p. 157)

The virtuous person will possess a set of dispositions which will include what Amelie Rorty calls *dispositions of interpretation, tropic dispositions, and self-activating dispositions*. (Rorty, 1988, pp. 137-8) Dispositions of interpretation ‘structure patterns of salience and importance: they organize the dominant and proper descriptions of situations.’ (Rorty, 1988, p. 137) The virtuous person is disposed to pick out the relevant or salient features of a situation, even when others fail to see such features as salient, or choose to focus on other aspects of the situation, rather than, without being prompted to, being ‘perceptively and interpretively sensitive to needs’ as, for example, the generous person will be. (*Ibid.*) The helpful person will see a person holding a map, looking bewildered, and recognise not only that he is lost, but also that he needs help. (Cf. McDowell, 1997) A helpful person must, as McDowell (1997, p. 142) puts it, ‘know what it is like to be confronted with a requirement’ of helpfulness.

Tropic dispositions ‘lead a person to gravitate to the sorts of situations that predictably elicit prized character traits.’ (Rorty, 1988, p. 137) So the helpful person would seek out opportunities to help. An example might be stopping to help someone gather their scattered papers when their briefcase has sprung open, in spite of being in a hurry, and already being late for a meeting. Self-activating dispositions ‘promote or create the occasions that require their exercise.’ (Rorty, 1988, p. 138)

Goldie (2000, p. 159) argues that the virtuous person will also possess *self-monitoring dispositions*. These are dispositions that ‘lead the person to
monitor her own responses as being in line with how, according to the trait, she thinks she ought to respond.’ (Goldie, 2000, p. 158) For example, Alice wants to go to a party but her mother needs some help around the house. Alice values being considerate and wants to act accordingly, so she asks herself whether going out would really be what the considerate person would do. (Cf. Goldie, 2000, p. 159)

It could be argued that some of the virtues ought to be exercised unreflectively. (Cf. Goldie, 2000, p. 159) Perhaps we should not need to reflect on what the kind person would do—we should always act kindly without having to think about it. But, as Goldie rightly points out, this is far too idealistic. (Ibid.) For we often feel competing pulls on our motivation, and find ourselves having to ask (ourselves) whether what we want to do is really what the virtuous person would do. The important thing seems to be that we care about doing the kind or generous or compassionate thing to the extent that we want to act accordingly. And besides, it seems that we would still be acting out of virtue even if we have to ask ourselves what the virtuous person would do. It is because we are virtuous that we are able to reflect on our own responses, and think about how we ought to respond.104

6.2.3 Feelings

The virtuous person must also have the right feelings: he or she must act virtuously without having to rail against conflicting desires. The akratic or incontinent person may have conflicting feelings: he might feel fear and yet manage to stay at his post. He will have acted courageously but he will not

104 I will be arguing below (in §6.4.1) that part of the virtue of compassion, when properly ‘trained’ in the fully virtuous person, is having a spontaneous empathic response, so for the fully virtuous person there is in fact no question of not being engaged motivationally and with feeling, as well as cognitively. I will also be arguing that you can come at this from another direction: if we value being compassionate, and we realise that we ought to be, yet we fail to feel compassion when we think we ought to, we can stimulate our empathic responses via HLE, which will function as a top-down process.
be fully virtuous because he experienced a competing desire to do otherwise—he wanted to run away rather than stay at his post. In the fully virtuous person, however, there will be no such conflict. The fully generous person, for example, will be ‘undividedly’ generous: he will act generously—do the right thing—without an inclination to the contrary. (Annas, 2006, p. 517)\(^{105}\)

### 6.2.4 Practical Wisdom

Being disposed to act as the virtuous person would act—without contrary inclination—is not, however, by itself sufficient for virtue. For even if a person were always inclined to be compassionate, or generous, or just, there are occasions when acting out of this motive will be inappropriate. A child might possess the natural virtue of honesty, such that he never tells an untruth.\(^{106}\) But this is not sufficient for possessing the virtue of honesty, for honesty is not always the best policy. Think of the child embarrassing her father by pointing out the size of a stranger’s bottom. Adults can be tactless and indiscreet, of course, but they have no excuse, for they ought also to possess practical wisdom (*phronesis*). Practical wisdom is the capacity to know what the right thing to do is on a given occasion, and the right way to go about it, given that one is already inclined to do the right thing. Put more simply, it is ‘knowledge of what one should do.’ (Hursthouse, 1999, p. 59) Practical wisdom is gained through life-experience, and is essential for being fully virtuous. It is particularly needed when there are *prima facie* demands from competing virtues, for example when adjudicating between the demands of justice and those of compassion, as the following example shows.

---

\(^{105}\) Also cf. Foot (1997, pp. 171-174).

\(^{106}\) See Aristotle (1999, p. 98) for his discussion of natural virtue and prudence as necessary conditions for virtue.
6.2.4.1 Promotion

There can be times when the demands of compassion conflict with the demands of other virtues, for example where questions of fairness come into play in determining what the right thing to do is, all things considered. In the following case, taken from empirical research undertaken on the conflict between compassion and justice, we can see how empathy-induced compassion pulls in one direction, and considerations of fairness in another. (Batson, et al., 1995)

Imagine, for example, an employer who values fairness and who must decide which of two employees to promote: Employee A is clearly better qualified and more deserving, but the employer feels sorry for Employee B, whose mother recently died. (Batson, 1995, p. 1043)

The employer’s compassion for B, if he used this as his criterion for action, would lead to his acting unfairly towards A, who is the best qualified for the job. The virtuous employer would have to weigh up considerations of fairness with considerations of compassion in order to reach his all things considered judgement as to whom, in this situation, he ought to promote. The employer is faced with a dilemma, and experiences competing motivational pulls. There is something inherently right in his feeling both these pulls: he wants to be fair, but he also feels compassion for B, both of which we would expect of the virtuous person. The virtuous person, who is both fair and compassionate, requires practical wisdom to determine what the right thing to do is, all things considered.  

---

107 Slote’s account of empathy would be unable to deal with examples like this, in which two competing motives are present, viz. a motive to be fair/just and a motive to be compassionate. This is because, on Slote’s view, empathic concern is the primary moral motive, and the motive to be fair/just derives from it. See (ECE, pp. 94-103).
6.2.5 Skills and Abilities

The virtuous person, then, will possess the virtues, central to which is practical wisdom, and also a wide range of skills and abilities. One of the skills he will possess is empathy. LLE is a capacity, however, not a skill—it is innate and involuntary, insofar as it is activated automatically, and we have no control over its acquisition or improvement. (Battaly, 2011, p. 290) It can, however, as we saw in the foregoing chapters, be modulated. Nevertheless, it is not a skill, for to count as a skill, it must be capable of being voluntarily exercised, which LLE cannot be.

Skills, on the other hand, can be voluntarily exercised, and can be improved with effortful practice. (Battaly, 2011, p. 209) In §1.3.3.3, I argued that HLE is a morally neutral skill, and not a virtue. It is sometimes considered to be a moral virtue, but it is not a suitable candidate for being one: like a virtue, HLE can be voluntary; but, unlike a virtue, it can be used for good or ill. Additionally, unlike a virtue, we can choose not to exercise the skill of HLE when it is appropriate without this detracting from our having the skill, and we can be motivated to HLE for all sorts of reasons, not all of which are good, and yet still be said to possess the ability to HLE.

6.2.6 Summary

In this section I have provided an overview of virtue and the virtuous person. The virtuous person must have the right motive, one that is appropriate to the virtue. Her motive must stem from her virtuous character. The virtuous person will possess a set of dispositions, including dispositions of interpretation, tropic dispositions, self-activating dispositions and self-monitoring dispositions. She must also have the right feelings: she will act virtuously without having to overcome conflicting desires. She will possess practical wisdom, which is the capacity to know what the right thing
to do is on a given occasion, and the right way to go about it, given that she is already inclined to do the right thing. Finally, the virtuous person will possess the capacity for LLE. She will also possess the capacity to HLE, which can be a skill, but is not, however, a virtue.

6.3 The Virtue of Compassion

6.3.1 Blum’s Notion of Compassion

In this section I argue that Lawrence Blum provides an account of compassion that improves on the sentimentalist notion of this virtue as depicted in Chapters 2, 3 and 4. Blum (1994, p. 175) defines compassion as one among a number of altruistic virtues, including sympathy, considerateness, and helpfulness, and describes it as follows:108

Compassion is not a simple feeling-state but a complex emotional attitude toward another, characteristically involving imaginative dwelling on the condition of the other person, an active regard for his good, a view of him as a fellow human being, and emotional responses of a certain degree of intensity.

According to Blum (1980, p. 133), the conditions for compassion are: (1) apprehension of the other person’s situation ‘as involving weal and woe’; (2) an appropriate emotional response; (3) being disposed, and motivated, to act altruistically in response to another’s suffering, where possible; (4)

108 An influential contemporary account of compassion is Martha Nussbaum’s (2001). Nussbaum has put forward a cognitive account of compassion according to which it is an evaluative judgement, as opposed to an altruistic attitude, which it is according to the notion of compassion that I am advocating in this chapter. Nussbaum’s concept of compassion is derived from her more general cognitivist account of the nature of the emotions. Challenging Nussbaum’s cognitive notion, Roger Crisp (2008, p. 241) has argued in favour of what he calls a ‘non-cognitive’ view according to which compassion is a ‘basic human emotion’ like anger or fear. Goldie (2002) has put forward a similar view. John Deigh (1994; 2004) argues that Nussbaum’s conception only accounts for what he calls ‘moral compassion’, which involves the judgement that the other is suffering and that his distress warrants our help. Deigh argues that Nussbaum’s notion cannot account for what he calls ‘nonmoral’ compassion, which involves feeling another’s pain and being motivated to alleviate their suffering without making any kind of evaluative judgement. Blum’s view appears to incorporate these two elements of compassion, and so would seem to be immune to Deigh’s objection. I do not have space in the thesis to discuss these alternative accounts of compassion in any detail. See de Sousa (1987), Goldie (2000), Wollheim (1999) and Stocker (1992) for other criticisms of the cognitivist view of emotion.
imaginative dwelling on the condition of the other person, and (5) a concern for the other’s good.

6.3.1.1 Apprehension of Weal and Woe

Compassionate action requires at least two things: firstly, apprehension of the other person’s situation ‘as involving weal and woe’; secondly, acting altruistically in response to that weal and woe. (Blum, 1980, p. 133) Blum (ibid., p. 132) claims that ‘part of what characterizes a person as caring, compassionate, sympathetic, or concerned is that he is more likely than other persons to apprehend situations in terms of the weal and woe of others.’ It is the mark of the compassionate person that he or she sees that a person is suffering in a situation when it is not altogether obvious. They will see the other person’s situation through compassionate eyes, to use Blum’s phrase. (Ibid.)

Whereas it is easy for most people to recognise that someone is suffering, especially if it is obvious (for example, if they are getting beaten up in front of you) not everyone will appreciate the finer details in some situations in which a person is not obviously suffering or in need, or will take these as providing grounds for compassion. Blum’s examples include one in which a person is digging his car out of some deep snow. There are two ways in which we might apprehend his situation: (1) qua man who is digging his car out of the snow; (2) qua man who is digging his car out of the snow and is having a tough time of it. (Blum, 1980, p. 132) The compassionate person will apprehend the man’s situation in the second way.

This apprehension does not require a kind of special sensitivity or ability to perceive a person’s weal and woe that only incredibly perceptive people possess. On the contrary: Blum stresses that failing to apprehend a person’s
weal and woe is not necessarily a failure of perception in the non-compassionate person. Rather, the compassionate person will be disposed to see the situation in terms of the person’s weal and woe. He will pick those features out as salient, and as reasons for action. It is the other person’s weal and woe that the compassionate person will notice in the situation; the same cannot be said for the person not in possession of this virtue.

6.3.1.2 Appropriate Emotional Response

Compassion, on Blum’s view, has an irreducible affective dimension. (Blum, 1994, p. 174) It has as its object another person (or class or group of people) who is (are) in a negative condition, ‘suffering some harm, difficulty, danger (past, present, or future).’ (Ibid., p. 173) The negative condition has to be ‘relatively central to a person’s life and well-being, describable as pain, misery, hardship, suffering, affliction, and the like.’ (Blum, 1994, p. 174) Things of a less serious nature, such as being made late for a meeting by a delayed train, may have a negative impact on a person’s life, but are not enough to constitute real suffering, of the kind that would be detrimental to a person’s overall well-being. (Cf. Goldie, 2000, p. 214)

We can have compassion for a person in a particular situation as well as compassion for a person’s general overall condition. For example, we can have compassion for a blind person on account of his blindness, qua overall life-affecting condition, even though he has achieved great things in spite of it, is generally happy, and does not let his disability affect him. (Blum, 1994, p. 174) We can also have compassion for a person whose overall life-condition is very good, but whose current condition is bad: a wealthy father, for example, who has just found out that his son has leukaemia. And the object of our compassion does not need to be aware of his condition, or, if he is, to see himself as suffering. As Adam Smith observed (see §3.2.4.2), we
can have compassion for the poor wretch who ‘laughs and sings’, yet is ‘altogether insensible of his own misery.’ (TMS, p. 9, I.i.1.12)

Compassion is an altruistic attitude that involves intense and enduring ‘distress, sorrow, hopes, and desires’, not just a ‘passing reaction or twinges of feeling’. (Blum, 1994, p. 178) Whilst we can experience twinges of compassion, Blum’s notion sees this as inferior to compassion proper, as a more ‘superficial and episodic version of “having compassion”’. (Blum, 1994, p. 179) Compassion is not just a fleeting feeling. Its affective dimension involves feeling appropriate levels of concern, distress, sorrow, and so on, for another person on account of their suffering. The level of concern will be higher in compassion than in other altruistic attitudes like pity, well-wishing, sympathy and helpfulness. (Ibid.)

6.3.1.3 Compassion and Pity

Blum rightly argues that compassion should be distinguished from pity. Hume and Smith, as we saw in Chapters 2 and 3, consider pity to be equivalent to compassion.109 But, as Blum and several other commentators have noted, pity has tones of condescension, superiority and even contempt for the person who is suffering, and should therefore be distinguished, and kept separate, from compassion.110 According to Blum, the difference between pity and compassion comes down to the way we view the person in distress: having compassion for someone ‘involves a sense of shared humanity, of regarding the other as a fellow human being.’ (Blum, 1994, p. 177) We have compassion for someone qua fellow human being who, just like us or anyone else, is capable of suffering. Pity, on the other hand, implies

---

109 Hume and Smith also argued that pity can easily blur into contempt. (Treatise, p. 387; TMS, p. 144, III.3.18)
that we view the other person’s suffering as something that ‘defines that person as fundamentally different’ from ourselves. (Ibid., p. 178) We do not entertain what has happened to them as a possibility for ourselves; we wish to keep ourselves separate, superior, untouched by what is afflicting the objects of our pity.111 The same is not true of compassion.

6.3.1.4 Disposition to Help Others

Having compassion requires being disposed to help others. The compassionate person will, where possible, help another person if they are in need. ‘Where possible’ here is meant to signify that we can still have compassion when: (a) we cannot do anything to alleviate the suffering of the other person without serious detriment to ourselves; (b) we could not do anything at all to help the other person even if we wanted to (if they have terminal cancer, for example); (c) helping would be inappropriate (it would jeopardise the person’s autonomy, for example).

6.3.1.5 Imaginative Dwelling

The imaginative dwelling on the condition of the other person characteristically involved in compassion involves imaginatively reconstructing the sufferer’s situation. This requires ‘imagining what the other person, given his character, beliefs, and values, is undergoing, rather than what we ourselves would feel in that situation.’ (Blum, 1994, p. 176) The imaginative reconstruction need not be great, and does not require exceptional imaginative powers. But we must, at the very least, ‘dwell in

111 For example, members of higher Indian castes might pity the Dalits in the slums of Mumbai. They see the condition of the Dalits as a ‘given’, almost as if their condition ‘goes with the territory’. They see the Dalits’ appalling poverty and position in the caste system as something that marks the Dalits out as distinct from themselves, for they are members of higher castes. They do not see the Dalits’ condition as a position they would ever find themselves in. Having compassion for the Dalits, on the other hand, involves considering them from the viewpoint of a shared humanity, *sub specie humanitatis*, as it were. When we have compassion for the Dalit on account of his condition, we are ‘acutely aware of a gap between that unfortunate condition and a condition consisting in that person’s flourishing.’ (Blum, 1994, p. 178) We see them as another, fellow human being, who, like us, is capable of suffering.
[our] imagination on the fact that [the other] is distressed’ in order to feel compassion for his plight. (Blum, 1994, p. 176) So a minimal amount of imaginative representation is necessary for compassion.

The ‘imaginative dwelling’, as Blum describes it, is not HLE in either of my senses. It seems more akin to what I referred to in Chapter 1 (§1.3.3.1) as ‘imagining-how-it-is’. (Cf. Goldie, 2011, p. 306) Imagining-how-it-is is an imaginative process, directed towards the psychological states of others, which involves the imagination but does not involve imagining from another person’s perspective, whether that is the empathizer’s own, or the target’s. When imagining-how-it-is, the empathizer imagines what a certain situation will be like for a certain person (or thing) in general, and not from any personal perspective. (Ibid.) This is the process that Blum claims is the minimum required for compassion.

Blum rightly points out that HLE (in-his-shoes imagining) can enhance compassion, and can lead to it, but is not sufficient for it. His reason for claiming that in-his-shoes-imagining is not sufficient for compassion is, however, different from mine. According to Blum, compassion has to arise solely from our consideration of the other’s distress, and not our own. If it were to arise as a result of imagining how we would feel in the other’s circumstances, then it would have arisen because of our having dwelt on our own distress, and not that of the other, Blum argues. (Blum, 1994, p. 176) Whilst I agree with Blum that in-his-shoes-imagining is not sufficient for compassion, I have argued that this is because of empathy’s motivational neutrality. Blum does not mention HLE (imagining being the other person), so it is not clear whether this form of HLE would be sufficient for compassion, on his view. Nor does Blum mention LLE, although it is clear that this form of empathy would be insufficient for compassion, according to
Blum, because it does not involve any imaginative dwelling whatsoever. The omission of imagining being the other person and LLE from Blum’s account of compassion does not detract from it as an account of compassion. But, as I argue in section 6.4, both these processes can enhance the virtue.

6.3.1.6 Concern for the Other’s Good

Finally, Blum’s notion of compassion requires that we have an active regard for the other person’s good. In other words, we must be concerned for their well-being. This will, according to Blum, be expressed in the manner in which we dwell on the other person’s plight. If my friend’s son falls ill, and I am compassionate, I will experience all sorts of thoughts, feelings and emotions, in addition to the belief that my friend and her son are suffering. For example: I will wish my friend’s son had not fallen ill; I will desire that he get better; I will have feelings of sorrow for my friend and her son, and I will think about what I could do to help. So possession of (the virtue of) compassion requires both a general concern for the well-being of the other person qua fellow human-being, and also a concern for the particular individual involved.

Hume, Smith and Slote also consider concern to be part of compassion. However, they argue that empathizing is sufficient for this concern, and that concern is a component part of compassion, rather than being a prerequisite for it. According to these sentimentalists, therefore, if we empathize, we will have compassion for, and be concerned for the well-being of, the other. For Hume (Treatise, p. 369), for example, as we saw in §2.3.1, compassion is a ‘concern for [...] the misery of others, without any friendship [...] to occasion this concern,’ and sympathy is sufficient for compassion. And according to Slote, compassion is a manifestation of empathic concern, which includes a concern for the well-being of the other person (MS, p. 76), and empathy.
mechanisms produce empathic concern. (See §4.2) However, empathy cannot be sufficient for concern, for empathizing will not guarantee that we act in the best interests of the target of the empathy, as the following example shows. Acting in another person’s best interests—with their well-being in mind—will sometimes require acting against our empathic reactions.

6.3.1.6.1 Drug Addict

A mother has a son who is a heroin addict. He comes to her for help because he is experiencing chronic withdrawal and needs his next fix, but has run out of money. He pleads with his mother to help him by giving him the money for a fix. LLE is likely to lead to the mother feeling empathic concern for her son, and being motivated to alleviate his suffering. HLE might also lead to her feeling empathic concern and wanting to help her son: he might implore her to imagine how he feels in his state of chronic withdrawal, or to imagine how she would feel if she were in his shoes. The deliverances of the mother’s empathy are likely to motivate her to alleviate his suffering; what would alleviate his occurrent suffering is a fix.

However, empathizing seems insufficient for concern in this case because, although giving her son drugs might well alleviate his short-term suffering, for it would doubtless ameliorate his immediate condition, it is not obvious that this would be in his overall, long-term interest. What is in his overall, long-term interest is treatment and rehabilitation. If the mother helped her son by giving him a fix, she could end up doing him more harm than good. Her empathy-motivated action could have negative consequences. Getting her son into rehab would likely require that the mother acts against her affective empathic responses by withholding money for a fix from her son. So it would seem that empathizing would not, in this case, guarantee a concern for her son’s overall well-being.
Blum’s notion of compassion captures the idea that, in order to be in possession of the virtue of compassion, the virtuous person must already have a concern for the well-being of the other person. The sentimentalist definitions of compassion, according to which empathizing is sufficient for concern and compassion, seem flawed, for empathizing is not sufficient for concern. It is perplexing that Slote, who partly models his notion of empathic concern on work by Batson (1991), claims that empathy is sufficient for concern. As we saw in §4.2, Slote endorses Batson’s empathy-altruism hypothesis, according to which empathy produces and sustains altruistic behaviour towards others according to their perceived need. (Batson, 1991; MS, p. 16) What Batson means by ‘empathy’ in his empathy-altruism hypothesis is empathic concern. For Batson, empathic concern includes emotions like sympathy, compassion and tenderness. All these are what he calls ‘other-oriented’ emotions that are ‘elicited by and congruent with the perceived welfare of another in need.’ (Batson, 2007) According to Batson himself, a concern for the good of the other person is what he calls a ‘situational antecedent’ for empathic concern. (Batson & Eklund, 2007, p. 65) That is, we must have a concern for another’s well-being in order for empathy to result in empathic concern. So Batson’s notion of empathic concern presupposes that we value the welfare of the person in need, as Blum’s does. (Batson, 2007, p. 65) It is because Blum’s notion of compassion presupposes concern, rather than including it, that I consider his account of compassion to be superior to the sentimentalist alternative.

6.3.2 Summary

In this section I outlined Blum’s notion of compassion. The necessary conditions for compassion are, on Blum’s view: (1) apprehension of the other person’s situation as involving weal and woe; (2) an appropriate emotional response; (3) being disposed, and motivated, to act altruistically in response to another’s suffering, where possible; (4) imaginative dwelling on the
condition of the other person, and (5) a concern for the other’s good. I argued that Blum’s notion of compassion is superior to that put forward by the sentimentalists because it requires a concern for the good of the other person as a necessary condition for this virtue. I argued that the sentimentalist notions of compassion, according to which compassion incorporates, but is not preceded by, a concern for the other’s good, are inadequate because empathizing is not sufficient for concern, and therefore cannot be sufficient for compassion, if compassion includes concern. I also argued that Blum’s notion of compassion lacks a role for LLE and imagining being the other person, but that this does not detract from the overall efficacy of his account.

6.4 The Role of Empathy in the Virtue of Compassion

In this section I argue, using examples, that empathy can play a positive role in morality by enhancing compassion.

6.4.1 Examples

6.4.1.1 Subway

Dan is riding seated on a subway train. He sees other people standing. One of them is an elderly woman holding a heavy shopping bag. She looks to be in considerable discomfort. The compassionate thing to do is for Dan to offer the woman his seat. Compassion requires Dan to be able to pick up on the elderly woman’s suffering, for an apprehension of her weal and woe is, as I argued in §6.3.1.1, necessary for compassion. In §6.2.2 I mentioned Rorty’s notion of dispositions of interpretation. These ‘structure patterns of salience and importance: they organize the dominant and proper descriptions of situations.’ (Rorty, 1988, p. 137) Virtuous people are disposed to pick out the relevant or salient features of a situation, and are

112 This is adapted from an example provided by Blum (1980, p. 130). Also see Goldie (2000, p. 158) for a similar example.
‘perceptively and interpretively sensitive to needs.’ (Ibid.) In this example, LLE could enhance Dan’s compassion by disposing him to pick out the relevant or salient features of the woman’s situation—her weal and woe. Dan’s LLE could also motivate him to help the woman by causing him to have an affective response, another necessary condition of compassion.

In previous chapters, I argued that our empathic responses can be modulated by various factors, including negative and prejudicial beliefs about others, which can lead to a failure to feel empathic concern. For example, in Chapter 4 (§4.4.3) I argued that we feel less as a result of empathizing with AIDS sufferers if we believe they have ‘brought it on themselves’. Given that Dan is virtuous, however, his virtue will rule out his holding such unwarranted prejudices. A person’s suffering as a result of having AIDS is always bad, and always warrants compassion, the virtuous person will think, regardless of how it was contracted. Because he is virtuous, Dan’s LLE will not be modulated by unwarranted prejudices, and he will be more likely to feel empathic concern as a result of empathizing.

However, as we saw in Chapter 4 (§4.3.2), some empathic biases arise not as a result of our beliefs, but rather because of the implicit bias inherent in the capacity for empathy. For example, we tend to feel a stronger empathic reaction for members of our own race, and less for members of other races. Assuming Dan is white, and the elderly woman is black, it could be argued that Dan may not be as sensitive to her suffering, because his empathy will have been modulated by his implicit empathic bias, over which he has no control. LLE would still enable Dan to pick up on the woman’s suffering, however, even though his reaction may be less strong than it would be for an ingroup member. And once he has made the attribution of suffering, he will know, because he possesses self-monitoring dispositions (see §6.2.2), that he
ought to feel compassion for her to the same extent as for someone who is a member of his own race or ethnicity. The self-monitoring dispositions of the compassionate person lead him to monitor his responses to see whether they are in keeping with how, according to compassion, he thinks he should respond. (Cf. Goldie, 2000, p. 158) Upon realising that he has failed to be as affectively motivated as he ought to be by the woman’s suffering, Dan could enhance his weak LLE response by ‘up-regulating’ it from the top down.

Top-down processes, as we saw in §5.4.3, can affect bottom-up responses to another’s suffering by either dampening them down or intensifying them. There is another way in which top-down processes can affect bottom-up processes that was not mentioned in Chapter 5, but is highly relevant here: HLE itself can function as a top-down process. It can serve to generate responses to another person’s suffering ‘by means of imagination or anticipation of the other’s state’, usually when direct perception of another’s state via LLE, that would cause an empathic response from the bottom up, is not possible. (Singer & Lamm, 2009, p. 11) So, even though Dan’s LLE response to the elderly woman may be less strong if she is not an ingroup member, he will nevertheless perceive her suffering, albeit only weakly. And Dan, if he is virtuous, will value being compassionate, and realise that he ought to feel more compassion in this instance. He could therefore use his HLE to up-regulate or intensify his bottom-up affective response from the top down by putting himself in the woman’s shoes or imagining being her.

Dan could employ this top-down stimulation by engaging in deliberate HLE. That is, he could deliberately put himself in the elderly woman’s shoes or imagine being her. He could also do this spontaneously. In Chapter 1 (§1.3.3.2) I argued that HLE can be deliberate and spontaneous. Perhaps part
of the virtue of compassion, in the fully ‘trained’ virtuous person, is being disposed to spontaneously HLE when we see another person suffering. Dan’s LLE, no matter how weak, could therefore trigger his HLE, which could serve to up-regulate his initially weak affective response to the woman’s suffering, thereby ensuring that, for the fully virtuous person, there is in fact no question of not being engaged motivationally and with feeling, both of which are necessary for compassion.

I have used this example to argue that empathy can play a positive role in enhancing Dan’s compassion by enabling him to be more reliably able to pick up on the elderly woman’s suffering, apprehension of which is necessary for compassion. I have also argued that empathy can enhance compassion by enabling Dan to stimulate his affective response to the woman’s suffering, if the former is weak due to the inherent bias of empathy, by either deliberately or spontaneously engaging in HLE, thereby ensuring that he is more reliably able to feel the affective response required for compassion. (See §6.3.1.2)

Although the elderly lady is suffering in Subway, she is not experiencing profound suffering of the kind that might be more likely to elicit personal distress as a result of empathizing, and perhaps lead to avoidance or non-

---

113 A similar example of how HLE can be used to stimulate an affective response is as follows. Joan is listening to the radio during the 1984 famine in Ethiopia. An advertisement by a leading charity comes on, asking for donations to fund humanitarian aid. Joan has seen and heard so many adverts like this recently, and she has, in the past, made many donations to various charities. She feels utterly unmoved by this charity appeal, even though she knows she ought to make a donation. Joan is experiencing a form of compassion fatigue, defined in the Oxford English Dictionary (1989) as follows:

Apathy or indifference towards the suffering of others or to charitable causes acting on their behalf, typically attributed to numbingly frequent appeals for assistance, esp. donations; (hence) a diminishing public response to frequent charitable appeals.

The virtuous person would recognise, as Joan does, that she ought to feel compassion for the famine victims. However, unlike Joan, she will HLE in order to stimulate her affective response(s).
altruistic helping rather than compassion. In addition to the apprehension of suffering and a motivation to alleviate that suffering, compassion requires an *appropriate* affective response: compassion’s affective dimension, on Blum’s view, involves feeling appropriate levels of concern, distress, sorrow and so forth. (Blum, 1994, p. 179) But as we saw in *Junior Doctor* (1) (§5.4.3.3), it is possible to become over-aroused by the suffering of another person to the extent that we avoid helping them, or help them in order to relieve our own distress, both of which might prevent compassionate action. In the next example I argue that empathy can enhance compassion by providing an appropriate affective response, even in the face of the kind of empathic over-arousal that caused Tim to fail to provide compassionate care for his patient in *Junior Doctor* (1).

6.4.1.2 Junior Doctor (2)

Tim is a junior doctor doing a rotation in Accident and Emergency. One of Tim’s patients, Al, is a burns victim. He is in agony, screaming with pain almost constantly. Tim knows he has to turn Al over in bed, or Al will get bed sores. But even the slightest movement causes Al indescribable pain. Being near to Al causes Tim intense emotional distress (via LLE, perhaps). He cannot face seeing him on his rounds, even though he knows he must do so. Tim *must* attend to Al, for he must turn him over to prevent bedsores from developing.

The right (compassionate) thing for Tim to do is to care for Al compassionately, turning him over carefully, being mindful of his pain and suffering. Tim has the right motive, for he cares about doing a good job, and about being a good doctor, and he is concerned for the well-being of his patients. As we saw in the previous chapter (§5.4.2), consistent exposure to the suffering of others can cause high levels of vicarious traumatisation,
otherwise known as burnout or compassion fatigue, in doctors and other care practitioners. (Palm, et al., 2004) Compassion is valued in care-givers, but too much exposure to another’s suffering can lead to personal distress, which can serve to prevent not only compassionate care, but any caring at all, due to avoidance behaviour. It can also lead to alleviation of the other’s suffering out of a desire to alleviate one’s own distress, rather than out of concern for the victim. Such helping action cannot be called compassionate because it is not done with a concern for the well-being of the sufferer in mind. (See §6.3.1.5)

In *Junior Doctor* (1), Tim became empathically over-aroused because he was prone to feeling intense negative emotions, and because he was not good at self-regulating his emotions. We saw that fully-trained doctors’ extensive medical knowledge enables them to unconsciously down-regulate their empathic reactions, such that they can avoid empathic over-arousal. Given Tim’s relative inexperience, he will be unlikely to be able to self-regulate in this way. However, we saw earlier (§5.4.4) that deliberate emotion self-regulation is also possible. This can occur in the form of emotion suppression, for example. (Butler, et al., 2003) Over-exposure to another person’s suffering can lead to personal distress, whereas optimum exposure tends to lead to empathic concern, sympathy and compassion. (Eisenberg & Fabes, 1992) If Tim were able to adequately self-regulate by making a deliberate effort to suppress his negative emotions, he might avoid being over-aroused by Al’s suffering, which will be more likely to result in caring, rather than avoidance behaviour.

However, conscious emotion self-regulation tends to be costly for the empathizer, as we also saw in §5.4.4. It can lead, for example, to physiological consequences such as raised blood pressure, and social
consequences such as decreased rapport with, and mutual feelings of stress and anxiety between, the empathizer and the target. (Butler, et al., 2003) In *Junior Doctor* (1), I concluded that empathy is not sufficient for caring because making the effort to suppress one’s emotions in order to care in the face of great suffering requires being willing to bear the cost of emotion self-regulation. Perhaps, if Tim in *Junior Doctor* (2) is compassionate, he will be prepared to bear the cost of self-regulation out of concern for Al’s well-being, a necessary precondition for this virtue.

In this example I have argued that empathy could enhance Tim’s compassion by enabling him to have an appropriate (optimum) affective response to the suffering of his patient, even in the face of intense suffering. This is because, if he is compassionate, he will be concerned enough for his patient’s well-being to bear the cost of the emotion self-regulation that can enable him to have the optimum affective response required for compassionate caring. In the next example I argue that empathy can enhance compassionate action by facilitating the requisite sensitivity to another’s feelings.

6.4.1.3 Jean

Blum (1980; 1994) argues that compassionate action requires knowledge and understanding. This knowledge and understanding can incorporate a sensitivity to, and awareness of, the feelings of the other person. Empathy can enhance such sensitivity. For example, consider Jean, an elderly woman being cared for by her daughter, Sandra, in Sandra’s home. Jean is sitting in the lounge with her son, Sandra’s brother, and his wife and children. Jean has had an accident—she has soiled herself—but is as yet unaware of this fact. If she were to notice, she would be humiliated, embarrassed and

---

ashamed. For not only is she a proud person who finds it difficult to accept that she needs help; she is also sitting in the lounge with her visitors, and will assume that they have known about her accident all this time but have been too polite to say anything.

Sandra realises that Jean has had an accident, and that she (Jean) has not noticed. The compassionate thing for Sandra to do is to help Jean, whilst being mindful of her feelings. This will require a sensitivity to Jean’s actual feelings (she is unaware of having had an accident) as well as her potential feelings (humiliation and shame). Sandra needs, for example, to change Jean’s clothes, but without her knowledge, so Jean will not become aware of having had an accident. Perhaps Sandra might suggest that her brother show the family round the house, or distract the children, so she can find a way to change Jean without her knowing. Or perhaps she could create some pretext for whisking Jean away momentarily. Whatever Sandra chooses to do, compassionate action requires her to help Jean in the right way, that is, with a sensitivity to her feelings. For if Sandra helps Jean in a way that is not mindful of her feelings, for example in a way that makes it obvious to Jean that she has had an accident, Jean is likely to suffer from the humiliation arising from the knowledge that she has been sitting there, in front of her family, in soiled clothing. And this will not be the compassionate thing to do.

The manner in which Sandra helps Jean is important, and empathy can make a difference. Because Jean is not currently suffering, Sandra’s LLE is unlikely to be triggered. But Sandra could spontaneously HLE with Jean, either by imagining being Jean or putting herself in Jean’s shoes. Both forms of HLE can result in Sandra coming to know how Jean will feel if she is made aware of her situation. However, as I argued in §3.3.1, in-his-shoes-imagining is
not sufficient for compassion because it can give rise to incorrect attributions of mental states to others: the phenomenon of quarantine failure can result in attributors’ own mental states seeping into their simulation routines. If acting compassionately depends on Sandra making a correct prediction about how Jean will feel if she discovers her accident, then HLE may not be able to enhance her compassion after all. Against this objection, we can say that, perhaps part of the virtue of compassion, in the fully trained virtuous person, is being able to make reliably accurate predictions by being skilled mindreaders.

More plausibly, it is possible to suggest that Sandra spontaneously engages in the other form of HLE, and imagines being Jean. In §5.5.1, I argued that a necessary condition for this form of empathy is an adequate characterization of the other person. (Cf. Goldie, 2000, p. 195) Given that Sandra is Jean’s daughter, we can assume that she knows her well, and is therefore able to imagine being her relatively easily. In addition, if Sandra is virtuous, there will be less danger of her engaging in HLE with a bad motive, as we saw is possible in §5.5.

Finally, these forms of HLE can combine, which could serve to enhance Sandra’s sensitivity to her mother’s feelings. Martin Hoffman (2011, p. 233) claims that what he calls ‘combined self/other focused perspective-taking’ can occur sequentially or simultaneously when we focus on another’s distress. I argued earlier (§6.3.1.5) that only minimal imaginative dwelling on another’s distress is necessary for compassion. (Cf. Blum, 1994, p. 176) As a compassionate person, perhaps Sandra need only dwell on the fact that her mother has soiled herself in order to then spontaneously engage in both forms of HLE with her, either sequentially or simultaneously.
In this example I have argued that empathy can enhance Sandra’s compassion by enabling her to be sensitive to her mother’s feelings. Sandra could spontaneously HLE with Jean, either by imagining being Jean or by putting herself in Jean’s shoes, both of which could result from minimal imaginative dwelling on her plight. Given that a sensitivity to the feelings of others is necessary for compassionate action, empathy can serve to enhance such action by improving this sensitivity.

6.4.2 Summary

In this section I have used examples to argue that empathy can play a positive role in morality by enhancing compassion. In Subway, I argued that empathy can play a positive role in enhancing Dan’s compassion by enabling him to be more reliably able to pick up on the elderly woman’s suffering. I also argued that HLE can enable Dan to stimulate his affective response to the woman’s suffering if the former is weak due to the inherent bias of empathy. In Junior Doctor (2), I argued that empathy could serve to enhance Tim’s compassion by enabling him to have an appropriate affective response to the suffering of his patient, even when this suffering is intense: if he is compassionate, he will be concerned enough for his patient’s well-being to bear the cost of the emotion self-regulation that can enable him to have the optimum affective response required for compassionate caring. In the final example, Jean, I argued that empathy can enhance compassionate action by enabling Sandra to be sensitive to her mother’s feelings, which allows her to help her mother in the manner appropriate to the virtue of compassion.

6.5 Conclusion

This chapter defended a particular interpretation of an instrumental view of empathy’s moral value. I argued that empathy can be a good thing, in the sense of being reliably for the good, when it is integrated in the right way with virtue. I argued that one such virtue with which it can be integrated is
compassion: in the hands of the virtuous person, empathy can play a positive role in morality by enhancing this virtue.

In section 6.2, I provided an overview of virtue and the virtuous person. I argued that the virtuous person will have the right motive, that is, the motive appropriate to the virtue, and that this must stem from her virtuous character. I argued that she will possess a set of dispositions, be practically wise, and have the right feelings. Finally, I argued that the virtuous person will possess the capacity for LLE, as well as the capacity to HLE, which I argued is a skill, but not a virtue.

In section 6.3 I described Blum’s notion of compassion, and argued that it is preferable as an account of this virtue to the sentimentalist notions advocated by Hume, Smith and Slote because it requires a concern for the good of the other person as a necessary precondition, rather than incorporating it, as the sentimentalist notions do. I also argued that Blum’s account does not include a role for LLE and imagining being the other person.

Finally, in section 6.4, I argued that all three empathy processes (in-his-shoes-imagining, imagining being the other person, and LLE) can enhance compassion. Using examples, I argued that empathy can play a positive role in morality by enhancing compassion. It can enhance compassion by enabling the compassionate person to pick up on the suffering of others more reliably, as well as enabling him to stimulate his affective response to the sufferer if this turns out to be weak, due to empathy’s inherent bias. I argued that empathy can enhance compassion by enabling the empathizer to have an appropriate affective response to the intense suffering of those in his
care: if he is compassionate, the empathizer will be sufficiently concerned for his patient’s good to bear the cost of the necessary emotion self-regulation. Lastly, I argued that empathy can enhance compassionate action by enabling the compassionate person to be sensitive to the other’s feelings, which in turn allows him to help the other in the manner appropriate to this virtue.
7 Conclusion

This thesis has argued that two ways of considering empathy’s role in ethics are fundamentally flawed, for they fail to take into account the myriad ways in which empathy can be affected and influenced by our motivations. Chapter 1 provided an overview of the thesis and set out my main objection—the motivation objection—to constitutive and instrumental views of empathy’s role in morality. The motivation objection is the general claim that empathy is motivationally neutral: we might empathize for morally good motives, or for morally questionable motives, or out of some other kind of motive, such as self-interest, which may or may not coincide with what is morally appropriate. Moreover, the emotions or behaviour that can result from the empathizing can be affected by the motive behind it, with the result that we can have an inappropriate emotional response or act inappropriately as a result of empathy.

The motivation objection has three aspects: (1) reliability; (2) function, and (3) circularity. Aspect (1) is the claim that empathizing does not always lead to the right results because it is motivationally neutral. Aspect (2) is the claim that, because of this motivational neutrality, empathy is not sufficient for various functions. And aspect (3) is the claim that its motivational neutrality means that empathy cannot be used to define or explain certain aspects of morality in a non-circular way. Chapter 1 introduced the empathy concept and described some of the phenomena that have been called empathy, many of which I referred to throughout the thesis, as well as explaining and justifying the distinction I have been making between low- and high-level empathy phenomena, according to whether or not they involve the imagination. I argued that empathy is a process, not an outcome, and introduced the notion of empathy as it appears in relation to the mental
simulation debate in the philosophy of mind. I identified two types of high-
level empathy—in-his-shoes imagining and imagining being the other
person—and argued that high-level empathy can be spontaneous and
deliberate. I also argued that high-level empathy is a capacity or skill, and
not a moral virtue.

In Chapters 2, 3 and 4, I applied the motivation objection to three
constitutive views of empathy’s role in morality. Chapter 2, as well as
providing a detailed exegesis of Hume’s concept of sympathy, applied
aspect (2) of the motivation objection to sympathy’s role in Hume’s moral
theory, arguing that imaginative sympathy, as I called Hume’s limited form
of high-level empathy, although necessary for compassion, is not sufficient
for it. Imaginative sympathy, I argued, cannot guarantee that compassion is
produced when we sympathize because what we feel as a result of
sympathizing depends on the disposition and character of the sympathizer.
Aspect (2) of the motivation objection was also applied to Hume’s theory of
moral judgement. I argued that a sympathizer’s motives mean that his
sympathizing cannot guarantee that he makes genuine moral judgements.
In addition, I argued that Hume’s theory cannot provide us with an
adequate explanation of why we should privilege the moral sentiments we
feel from the common point of view of morality and act in accordance with
these, rather than our situated sentiments. I also introduced aspect (3) of the
motivation objection in this chapter, even though Hume does not face it
directly, in order to pave the way for its use in Chapters 3 and 4. I argued
that, for Hume, sympathy is necessary for knowing virtue and being
virtuous, and the outputs of sympathy are dependent on having the right
character and disposition, which means that Hume’s view appears to be
circular: sympathy appears to presuppose virtue, for sympathizers must
already possess the right disposition and character—which is tantamount to being virtuous—in order to know virtue and be virtuous.

In Chapter 3 I introduced the second of the three constitutive views. I argued that Smith faces all three aspects of the motivation objection, and defended the view that he is a proto-simulationist. I applied aspects (1) and (2) to Smith’s claim that high-level empathy is sufficient for compassion. He faces aspect (1), I argued, because high-level empathy processes are unreliable, frequently giving rise to incorrect attributions of mental states to others. I also argued that the sympathizer’s disposition and character can affect the outcome of his high-level empathy, and concluded that it is therefore insufficient for compassion. I argued that aspect (2) applies to Smith’s theory of moral judgement: sympathy is not sufficient for moral judgement because a sympathizer’s disposition and character can cause him to falsely imagine how he would feel in someone else’s shoes, potentially resulting in incorrect and partial moral judgements. Smith could invoke the impartial spectator to ensure partiality and the elimination of subjective bias in moral judgement, but I argued that this would require spectators to have a motive to adopt his perspective and seek his approval. Spectators, I argued, want to obtain this approval because doing so means they are acting virtuously, which they in turn desire to do because they value virtue. Finally, I argued that Smith faces aspect (3) of the motivation objection because his explanation of virtue in terms of sympathy appears to be circular. He claims that virtue can be explained in terms of seeking sympathy with the impartial spectator. However, I argued that sympathizing with the impartial spectator requires being virtuous, so virtue cannot be explained in terms of sympathy in a non-circular way.
In Chapter 4, I applied aspects (2) and (3) of the motivation objection to a contemporary sentimentalist view with a constitutive role for empathy. Michael Slote claims that empathy is sufficient for moral obligation. However, I argued that, because empathy is inherently racially biased, and we therefore feel more empathic concern for members of our own race, basing moral obligation on the strength or weakness of our empathic reactions could lead to justification of preferential helping along racial lines. I also applied this aspect, aspect (2), to Slote’s claim that empathy is sufficient for the moral sentiments. He claims that empathizing is sufficient to cause us to feel approbation and disapprobation when others display or fail to display empathic concern in their actions. However, I argued that empathy can be modulated by our beliefs and prejudices about the person doing the action, which can in turn affect the outcome of the empathizing. Finally, I argued that Slote faces aspect (3) of the motivation objection. He claims that empathic concern is the fundamental moral motive, and that acting virtuously consists in acting with this motive. He also claims that empathy mechanisms cause empathic concern. Virtue is therefore defined as empathic concern, which is a product of empathy mechanisms. I argued that we will only be guaranteed to feel empathic concern for others as a result of empathizing if we are not already prejudiced towards them, and if we do not stigmatize them. If not being prejudiced is characteristic of virtue, then empathy mechanisms will only produce empathic concern if the empathizer is already virtuous. Slote’s definition of virtue in terms of empathic concern therefore looks circular, for if virtue is empathic concern, and feeling empathic concern requires virtue, then virtue cannot be defined as empathic concern without presupposing what it is trying to define, namely virtue.

In Chapter 5, I applied aspect (2) of the motivation objection to three instrumental views of empathy’s role in morality. I argued that neither form
of high-level empathy is sufficient for moral reasoning because an empathizer’s motives can lead to his forming the wrong conclusion as to his moral obligations, and empathizing cannot tell the empathizer whose perspective he ought to attach more importance to: his own or the perspective of the person with whom he is empathizing. I also applied aspect (2) to Simon Baron-Cohen’s claim that empathy is of instrumental value in resolving conflicts. I argued that empathizing is not necessarily an aid to conflict resolution because we can form mental barriers that prevent our empathizing with those we consider to be immoral. We can overcome these barriers if we try hard enough, but the phenomenon of motivated inaccuracy can lead to empathizers making inaccurate predictions of others’ mental states, which can result in a failure to resolve the conflict in question. In addition, I argued that empathy is not sufficient for caring because our motivations can affect how, and whether, we will help or care for another when they are suffering or in need. Overexposure to prolonged and intense suffering can lead to empathic over-arousal, which can result in a failure to care. Empathizers can regulate their emotions by suppressing them, and so can care as a result of empathizing, but this can be costly, and so would seem to require that the empathizer be prepared to bear the cost, which would in turn require that she already cares. I also argued that the motivational neutrality of empathy opens it up to abuse: it can be engaged in with a bad motive; it can affect the manner in which the empathizing is undertaken, and it can affect the imaginative processes that can subsequently arise.

The conclusions reached in Chapters 2, 3, 4 and 5 were largely negative—they can be taken as a cautionary tale against empathy’s moral value. However, in Chapter 6 I argued that empathy can still have an important role to play in ethics. I argued that, when properly integrated with virtue, in particular the virtue of compassion, empathy can play an important,
instrumentally valuable role in our moral lives. It can be a good thing, in the sense of being reliably for the good, when it is integrated in the right way with virtue. In the hands of the virtuous person, empathy can play a positive role in morality by enhancing this virtue. Indeed, all three forms of empathy (in-his-shoes-imagining, imagining being the other person, and LLE) can enhance compassion. I argued, using examples, that empathy can enhance compassion by enabling the compassionate person to pick up on the suffering of others more reliably, as well as enabling him to stimulate his affective responses if these are weak due to empathic bias. Empathy can also enhance compassion, I argued, by enabling the empathizer to have an appropriate affective response to the intense suffering of those in his care. For, if he is compassionate, the empathizer will be prepared to shoulder the possible negative consequences of emotion self-regulation. Finally, I argued that empathy can enhance compassionate action by enabling the compassionate person to be sensitive to the feelings of the other person, enabling him to help in the manner appropriate to this virtue. Chapter 6 also served to justify the circularity aspect of the motivation objection that I applied to the sentimentalist views, for virtue must precede empathy in order for empathizing to be reliably for the good.
Bibliography


Cheng, Y. et al., 2007. Expertise modulates the perception of pain in others. *Current Biology*, 17(19), pp. 1708-1713.


