The Ethics and Epistemology of Empathy

A dissertation presented

by

Olivia Bailey

to

The Harvard University Department of Philosophy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Philosophy

Harvard University

Cambridge, Massachusetts

April 2018
The Ethics and Epistemology of Empathy

Abstract

Empathy is a familiar form of emotionally charged imaginative perspective taking. In this dissertation I offer an account of empathy’s moral importance that emphasizes the special value of its unique epistemic functions. Specifically, I defend what I call the humane understanding thesis: empathy is the source of a distinct epistemic good, humane understanding, which consists in the appreciation of the intelligibility of others’ emotional perceptions, and humane understanding is necessary for (and, indeed, partially constitutive of) fully virtuous relations with other people.

Adam Smith held that empathy is the unique means by which we judge the propriety of other’s emotions. I draw out the kernel of truth in this surprising claim. Emotions inevitably present the world in a certain evaluative light. When I see my situation as calling out for my anger, or fear, or adoration, my emotion appears to me to be properly responsive to my situation. I can judge my emotion to be inappropriate, all things considered. However, I cannot both be angry at your betrayal and be totally mystified by my own anger, or resent you but see no absolutely nothing in your conduct that appears to call out for my resentment. This insight applies to empathy. When we empathize, we feel genuine emotions. And when we feel these emotions, we cannot fail to appreciate the intelligibility of the other’s original emotion first-hand. This appreciation constitutes humane understanding of the other’s emotion. I further clarify the nature of this epistemic achievement by distinguishing it from two other nearby forms of understanding, one that is achieved through
narrative rather than perspective taking, and one that is achieved through unemotional perspective taking.

Empathy has recently come in for some heavy criticism. It has been described as not only morally unnecessary, but positively morally corrosive. Recognizing that empathy is the unique means by which we secure humane understanding of others’ emotional perceptions (and, by extension, the choices and actions that stem from those perceptions) enables us to appreciate aspects of empathy’s moral significance that its most prominent detractors have neglected. Humane understanding is a powerful way of being close to another person, and closeness is something that social beings like ourselves crave, not for any further or deeper purpose, but just because it is valuable to us in and of itself. People have a powerful need to be empathized with because empathy relieves a painful form of loneliness. Empathy can be a way of caring, rather than just a means of getting us to care.

In further defense of empathy’s moral significance, I also argue that empathy has a meaningful role to play in the development of both our reflective beliefs about moral properties and our emotional sensitivity to those same properties. Empathy can allow us to discover new moral distinctions through imaginative emotional experience, and it can also help us to shape and sharpen our moral emotional perception. Finally, I round out my review of empathy’s moral significance with an argument that empathy with others makes it very difficult to intend to treat them in certain immoral ways. We cannot both empathize with someone and see them as unfeeling, as unthinking, or as lacking a normative perspective on the world. Consequently, empathizing with someone is at least not compatible with treating them as a mere object.

I offer my Smithian picture of empathy’s moral and epistemic functions as an alternative to the popular Humean view that empathy moves us from our default position of solipsism and egoism
to knowledge of and care for other’ inner lives. That view’s ambition is tantalizing, but it distorts both the nature of empathy and the nature of our other-oriented concern and understanding. Hume’s account of empathy’s significance, and the more recent accounts that take after it, end up making it look as though empathy simply contributes to the expansion of egoism, and/or as though empathy depends for its motivational efficacy on a kind of epistemic mistake.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1. SHELLEY’S “GREAT INSTRUMENT”</td>
<td>1</td>
</tr>
<tr>
<td>2. A BRIEF OVERVIEW</td>
<td>3</td>
</tr>
<tr>
<td>CHAPTER ONE: A PORTRAIT OF EMPATHY</td>
<td>11</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>2. EMPATHY’S ESSENTIAL FEATURES</td>
<td>12</td>
</tr>
<tr>
<td>2.1. IMAGINATION</td>
<td>12</td>
</tr>
<tr>
<td>2.2. GETTING AT OTHERS’ EXPERIENCES FROM THE INSIDE</td>
<td>18</td>
</tr>
<tr>
<td>2.3. THE ROLE OF EMOTION</td>
<td>27</td>
</tr>
<tr>
<td>2.4. EMOTIONAL CONTAGION</td>
<td>33</td>
</tr>
<tr>
<td>3. HOW WE THINK EMPATHY MATTERS: THE BRIDGING PICTURE</td>
<td>37</td>
</tr>
<tr>
<td>4. DESIDERATA FOR AN ACCOUNT OF EMPATHY’S NATURE AND SIGNIFICANCE</td>
<td>49</td>
</tr>
<tr>
<td>5. FROM THE PRELIMINARY PORTRAIT TO AN ENLIGHTENMENT CONCEPTION</td>
<td>51</td>
</tr>
<tr>
<td>CHAPTER TWO: HUMEAN EMPATHY: AN IDEA AND ITS AFTERLIFE</td>
<td>53</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>53</td>
</tr>
<tr>
<td>2. EGOCENTRIC PRIMACY IN THE TREATISE</td>
<td>56</td>
</tr>
<tr>
<td>2.1. HUME’S PHILOSOPHY OF MIND: THE BASICS</td>
<td>57</td>
</tr>
<tr>
<td>2.2. THE NATURE OF THE PASSIONS</td>
<td>59</td>
</tr>
<tr>
<td>2.3. HUME’S EGOCENTRIC PRIMACY</td>
<td>60</td>
</tr>
<tr>
<td>3. THE TWO-PHASE OPERATION OF SYMPATHY</td>
<td>65</td>
</tr>
<tr>
<td>4. THE JOY OF SYMPATHIZING</td>
<td>71</td>
</tr>
<tr>
<td>5. THE “IDENTICAL MOTIVATION” ACCOUNT</td>
<td>75</td>
</tr>
<tr>
<td>6. A “TRANSFORMATIVE MOTIVATION” ACCOUNT?</td>
<td>85</td>
</tr>
<tr>
<td>7. CONCLUSION</td>
<td>92</td>
</tr>
<tr>
<td>CHAPTER THREE: EMPATHY AND UNDERSTANDING</td>
<td>95</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>95</td>
</tr>
<tr>
<td>2. A SHIFT AWAY FROM HUME</td>
<td>98</td>
</tr>
<tr>
<td>3. AN INSTANCE OF SMITHIAN EMPATHY</td>
<td>104</td>
</tr>
<tr>
<td>3.1. SETTING THE IMAGINATIVE SCENE</td>
<td>105</td>
</tr>
<tr>
<td>3.2. EMPATHETIC EMOTION</td>
<td>107</td>
</tr>
<tr>
<td>3.3. THE MOMENT OF JUDGMENT?</td>
<td>114</td>
</tr>
<tr>
<td>4. TWO PROBLEMS FOR SMITH’S PROPOSAL</td>
<td>116</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.1. THE FIRST PROBLEM</td>
<td>117</td>
</tr>
<tr>
<td>4.2. THE SECOND PROBLEM</td>
<td>118</td>
</tr>
<tr>
<td>5. FROM JUDGMENT TO HUMANE UNDERSTANDING</td>
<td>125</td>
</tr>
<tr>
<td>6. THE NARRATIVIST CHALLENGE</td>
<td>136</td>
</tr>
<tr>
<td>6.1. SOME KIND OF THE HUMANE UNDERSTANDING THESIS</td>
<td>136</td>
</tr>
<tr>
<td>6.2. NARRATIVES AS THE KEY TO INTELLIGIBILITY?</td>
<td>139</td>
</tr>
<tr>
<td>7. THE CO-COGNITIVIST CHALLENGE</td>
<td>147</td>
</tr>
<tr>
<td>8. CONCLUSION</td>
<td>154</td>
</tr>
</tbody>
</table>

**CHAPTER FOUR: THE MORAL VALUE OF HUMANE UNDERSTANDING**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>157</td>
</tr>
<tr>
<td>2. AGAINST EMPATHY</td>
<td>160</td>
</tr>
<tr>
<td>3. THE NON-INSTRUMENTAL MORAL VALUE OF HUMANE UNDERSTANDING</td>
<td>166</td>
</tr>
<tr>
<td>3.1. EMPATHY AND HUMANE UNDERSTANDING IN A ROOM WITH A VIEW</td>
<td>166</td>
</tr>
<tr>
<td>3.2. VARIATIONS IN THE NEED FOR HUMANE UNDERSTANDING</td>
<td>174</td>
</tr>
<tr>
<td>3.3 HUMANE UNDERSTANDING AND MORAL HAZARD</td>
<td>180</td>
</tr>
<tr>
<td>3.4 HUMANE UNDERSTANDING AND FRIENDSHIP</td>
<td>183</td>
</tr>
<tr>
<td>4. EMPATHY AND MORAL EPISTEMOLOGY</td>
<td>187</td>
</tr>
<tr>
<td>5. EMPATHY AND MORAL MOTIVATION</td>
<td>196</td>
</tr>
<tr>
<td>6. CONCLUSION</td>
<td>205</td>
</tr>
</tbody>
</table>

**CONCLUSION**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. THE COURSE OF THE INVESTIGATION</td>
<td>208</td>
</tr>
<tr>
<td>2. HOW MY ACCOUNT MEASURES UP</td>
<td>211</td>
</tr>
<tr>
<td>2.1. THE FATE OF THE BRIDGING PICTURE</td>
<td>211</td>
</tr>
<tr>
<td>2.2. REVISITING THE DESIDERATA</td>
<td>214</td>
</tr>
<tr>
<td>3. WHAT REMAINS TO BE SAID</td>
<td>219</td>
</tr>
</tbody>
</table>

**BIBLIOGRAPHY**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>221</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

This dissertation has been a long time in the making, and as the years of research and writing have accumulated, so too have my debts of gratitude. It is a very great pleasure to be able to thank those who have helped me along the way.

First, I am grateful for the funding that enabled me to pursue this project. Funding from the Santayana Fund, the Martin Prize Fellowship, and the Safra Center for Ethics allowed me to take leave from teaching for two years and concentrate solely on putting my jumbled thoughts about humane understanding into order. I feel lucky to have received this extraordinarily generous support.

I became interested in empathy and in Adam Smith through the Philosophy and Literature Reading Group, which tackled The Theory of Moral Sentiments in my first year as a graduate student at Harvard. The time we spent puzzling over Smith's descriptions of guilt-wracked criminals, stoic heroes, and grieving mothers was one of the highlights of my time at Harvard, and I am thankful to the members of the group for those invaluable discussions. The members of the History of Philosophy Works in Progress Group and of the Moral and Political Philosophy Workshop have borne the burden of listening to me talk about Smith, Hume, empathy, and humane understanding for years, and I am also grateful to them for their patience and their insight. Some of the material in this dissertation has also benefitted from airings in the philosophy departments of Princeton University, the University of Chicago, the University of Aberdeen, Oxford University, the University of Pennsylvania, the University of Sheffield, the University of California, Berkeley, Wayne State University, the University of Pittsburgh, the University of Richmond, Queen’s University, Florida International University, Tulane University, and Loyola Marymount University. I am grateful to have
received many challenging questions, as well as much-needed encouragement, from the philosophers who attended those talks.

I am grateful to the entire faculty of the Harvard Department of Philosophy. Many of its members have contributed to this project through their brilliant lectures and thought-provoking questions. I have also benefitted from the aid of several philosophers outside of Harvard. Conversations with Roger Crisp, Jennifer Whiting, and Quassim Cassam have led me to rethink some of less promising ideas about empathy. I am particularly grateful to Samuel Fleischacker for our exchanges about The Theory of Moral Sentiments: I have taken great pleasure in the harmony of our sentiments concerning the TMS, just as Adam Smith would have predicted.

I owe so much to Harvard’s graduate students in philosophy. Doug Kremm and Zeynep Soysal were with me from the start, there for many late night conversations about big questions. James Bondarchuk and Sandy Diehl provided wonderfully helpful feedback on my work. The women of Emerson 306—Rachel Achs, Cat Wade, and Kate Vredenburgh—saw me through a difficult recovery from hip surgery in my final year at Harvard with endless good humor. I am grateful for their philosophical insight, and still more grateful for their friendship.

I am thankful to my family for their love, and to Alp Sipahigil for his unerring patience, thoughtfulness, and kindness.

It remains for me to thank the long-suffering members of my committee. Matt Boyle helped me to see that problems I was wrestling with were almost invariably deeper than I had thought. Susanna
Siegel’s relentless optimism led me to believe that the dissertation could in fact be written, and I have learned a great deal from her about good philosophical writing.

Finally, this dissertation is dedicated to Dick Moran. Dick saw me through many dark years, stubbornly believing in my abilities even when I did not. He is an exceptional mentor and a model of humane understanding, and I hope to continue to learn from him for many years to come.
INTRODUCTION

A man, to be greatly good, must imagine intensely and comprehensively; he must put himself in the place of another and of many others; the pains and pleasures of his species must become his own. The great instrument of moral good is the imagination.

Percy Shelley, A Defense of Poetry

1. Shelley’s “Great Instrument”

In the above passage from “A Defense of Poetry,” Shelley describes a perfected version of an activity that we his readers are supposed to intimately acquainted with, at least in a diminished or incomplete form. Both his characterization of this imaginative effort and his claim about its moral significance are presented not as the products of a chain of argumentation, but rather as a reminder of what we already know. And, indeed, Shelley’s statements here have the ring of familiar truths. We take ourselves to have extensive experience of putting ourselves in others’ places, and also of being the object of other’s perspective-shifting endeavors. This way of relating to others is one that we register as not only possible, but also quite normal. It similarly sounds right to say that a world in which no one related to anyone else in this way would be considerably impoverished. Of course, we think, it matters that we be able to walk in others’ shoes, to use our imaginations to get outside of our blinkered focus on our own weal and woe. That latter claim might look so obvious as to be uninteresting, even if it is perhaps mildly idiosyncratic of Shelley to deem the activity in question the great instrument of the good.

In addition to celebrating poetry for enhancing our ability to take on others’ pains and pleasures, Shelley also underlines its ability to make familiar phenomena “be as if they were not familiar.”

---

1 Shelley (1904): 34.

needn’t get poetical, though, to uncover the respects in which his familiar valorization of imaginatively adopting others’ pains can start to look mysterious. Ask one question, and others quickly follow: What does it actually mean to “make others’ pains and pleasures one’s own”? How could making others’ pains and pleasures our own possibly contribute to our relating to others in a morally good way? Should we trust our own sense that a lack of the abilities Shelley describes would consign one to moral mediocrity? Or should we instead think that there are other ways of being “greatly good,” and that this isn’t even a particularly promising path to virtue? The central aim of this dissertation, simply put, is to determine what these few lines from Shelley get right. To do so, I will need to take up all the aforementioned questions, and more besides.

I’m going to call the phenomenon Shelley describes “empathy.” “Empathy” is a newcomer to the English language, having been introduced in the 20th century as a translation of the German philosopher Robert Vischer’s 1873 neologism “Einfühlung” (literally, ‘in-feeling’), and it has proven to be handily versatile. In the philosophical and psychological literature, “empathy” has been used to refer to a surprisingly large variety of phenomena, some of which are very clearly not Shelley’s “great instrument.” To give just one example, some psychologists label infants’ tendency to join in when they hear crying an “empathetic response,” even though no one seriously thinks that a newborn’s crying response could be the result of imaginatively engaging with the experience of other wailing children. There is quite a lot of variation in the extra-academic use of “empathy”, as well. I choose to use the world “empathy” for the thing Shelley describes because some philosophers and many laypeople do also use it to pick out the phenomenon that he is concerned with, but my aim is not to provide an account of the nature and significance of everything that people call empathy.

Especially given the recent vintage of the word, it would be unwise to place too much stock in whether it strikes us as apposite to use “empathy” in some context or other. I will have things to say about how the thing I call empathy is related to other phenomena that travel under the same name, though.

2. A brief overview

My investigation of empathy’s nature and significance will unfold over the course of four chapters. I will begin in the first chapter by drawing upon Shelley’s statement to identify the key features of the target phenomenon. Part of the reason to lean on Shelley, here, is that his declaration resonates. It picks out a known phenomenon and offers a relatively familiar evaluation of its importance. What is more, though, there are ambiguities in Shelley’s remark that reflect some of the difficulties we will encounter in trying to describe empathy precisely. There are places where the clarity and univocality of our sense of the phenomenon gives way, and they are important. A failure to consider these points carefully will end up making empathy and our valuation of it look less interesting and less problematic than they really are.

My claim for the characterization of empathy I will articulate is not that most of us lack a more detailed sense of the phenomenon. It is rather that (1) the aspects of empathy I identify, at this level of generality, reflect powerful and widely shared intuitions, and (2) beyond this level of generality, we encounter substantial uncertainties. Here is the core characterization of empathy I find reflected in Shelley’s declaration: empathy is (1) an imagination-implicating mental activity that (2) involves encountering others’ experiences “from the inside” and that (3) somehow involves the emotional life of the empathizer. I review each of these characteristics. I indicate how they distinguish empathy from related psychological phenomena. And I consider just how much this basic description of
empathy leaves unsettled. I then sketch out an enduringly popular, broad vision of how empathy matters. From the Scottish Enlightenment through the contemporary resurgence of interest in the nexus of imagination, emotion, and morality, empathy has been seen as a bridge by which we may get outside of ourselves. The important differences between our concern for and understanding of ourselves, on the one hand, and our concern for and understanding of other people, on the other, suggest a problem to which empathy may be the solution. Empathy is cast as a or the means of escape from an egoism and solipsism that is natural to us. In order to guide my investigation of the idea that empathy could perform this bridging function, I conclude the first chapter by identifying desiderata for an adequate account of empathy’s moral significance.

In the second chapter, I turn to the earliest extended philosophical description of the way that empathy matters, David’s Hume’s account in A Treatise of Human Nature. Hume’s claims on behalf of empathy are appealingly ambitious. He proposes that when we empathize, we experience of copy of another’s feeling about her own suffering, or good fortune, or other circumstance. The empathetic process secures for us both knowledge of the other’s mind and non-instrumental concern for the other. Elements of this account are still very much present in the philosophical and psychological literature today, and their endurance is a testament to their intuitive appeal.

However, there is a deep problem at the heart of Hume’s account. The exact contours of the difficulty will take some time to ferret out, in part because Hume seems to offer two different explanations of how empathy generates care, but we can sum the problem up as follows. It looks like the complex mental mechanisms that Hume describes as collectively comprising the empathetic response don’t offer a route to genuine care for others. Instead, they generate a motivational state that depends on a kind of epistemic mistake, specifically a neglect or forgetting of the fact that the
other’s experiences and situation are the other’s, and not one’s own. Empathy therefore ends up seeming like a mental process that will only reinforce our self-care, or expand its scope.

After articulating this worry, I explore several responses to it on Hume’s behalf. I address the suggestion that empathetic emotion could be suitably transmogrified through the agent’s own awareness of its provenance (a suggestion that figures prominently in the developmental psychological literature on empathy), and I also consider whether a rethinking of our own relation to our own joy and suffering could rescue Hume’s approach without yielding a distorted picture of our psychology. Each of these paths meets with substantial problems. I propose that we instead adopt a new approach. We need to rethink the relationship between empathy’s epistemic and motivational functions. Hume treats empathy’s motivational significance as neatly separable from its epistemic significance. However, I propose that if we want to get a proper grip on why empathy matters morally, we have to begin with its epistemic import.

While it is common to mention a connection between empathy and knowledge or understanding, moral philosophers’ characterizations of the epistemic benefit of empathy and how it matters tend to be rather gestural. There is some sense that the epistemic good that empathy supplies is a knowledge or understanding that is somehow better than we could otherwise achieve. It is suggested that “the kind of information we can get from empathizing with someone is, at least sometimes, a richer, deeper kind of information,” that empathetic understanding is a “way of gaining a deeper understanding of what it is like,” and that empathy plays an essential role in “true understanding.”

Still, we are not often told what this “richness” or “depth” consists in, or how it might be ethically important. Part of the reason people don’t say much about these matters, I suspect, is that the

---

answers seem too obvious. The story about how empathy’s epistemic fruits matter morally looks like it should be easy to tell: We know that conflict and neglect arises from ignorance and misunderstanding, or their cousin, inattention. Empathy must furnish knowledge about how things are for another person in a way that implicates the emotions of the empathizer, and emotional thought is the kind of thing that orients attention particularly powerfully, so it follows that empathy will be a way of alleviating both problems of ignorance and problems of attention.

We ought not be content with this simple story. For one thing, it does not explain the significance of all of empathy’s distinctive features: one of empathy’s essential features is that it involves perspective taking, but the simple story doesn’t tell us anything about why that feature could matter. The simple account might as well be a story about any kind of perception, recognition, or knowledge of other’s experience that isn’t unemotional, or “cold.” In the third and fourth chapters of the dissertation, I develop a more satisfying picture of empathy’s epistemic significance, and I explain how empathy’s characteristic epistemic good matters morally. More specifically, I defend what I will call the humane understanding thesis: empathy is the source of a distinct epistemic good, humane understanding, which consists in the appreciation of the intelligibility of other’s emotional perceptions, and humane understanding is necessary for (and, indeed, partially constitutive of) fully virtuous relations with other people. To discover what makes empathetic understanding of another’s perspective and circumstances unique, and to appreciate why that understanding is morally significant, I return to the Scottish Enlightenment, and to Adam Smith’s *The Theory of Moral Sentiments.*

In Chapter Three, I draw on Smith’s work in order to defend the first half of the humane understanding thesis, the claim that empathy is the source of a distinct epistemic good. According to Smith, empathy is unique means by which we judge the propriety of others’ emotions, which Smith
rightly conceives of as evaluative perceptions (that is to say, perceptions of circumstances or entities as delightful, evil, frightening, joyous, hideous, and so forth). The idea that empathy is the sole means by which we come to approve or disapprove of others’ emotions was roundly criticized by Smith’s contemporaries, including Thomas Reid, and this criticism is not entirely misplaced. Smith’s way of thinking about the link between approval and empathy doesn’t consistently acknowledge the existence of non-empathetic paths to evaluative judgments. In making my empathetic feelings the standard by which I assess the propriety of your emotions, Smith also runs into the problem that I can and do sometimes judge my own emotions to reflect the evaluative characteristics of my situation inaccurately. There is a gap between experiencing an emotion and judging it to properly reflect reality, and so my own emotions will not always constitute an appropriate judgmental standard.

I argue that while Smith is wrong to posit such a strong connection between empathy and judgments of propriety, he is on the right track. When we manage to empathize with another, we are not bound to judge their emotions to be appropriate, but we are bound to appreciate the intelligibility of their emotions. An emotion is intelligible just in case it tracks the apparent evaluative qualities of its intentional object, and when we experience an emotion from the inside, as it were, it is not possible to regard that emotion as failing to track the evaluative qualities of its intentional object.

I call appreciation of the intelligibility of another’s emotion humane understanding because as with other forms of understanding, its possessor has an active and direct grasp of the relevant phenomena. When I empathize with your outrage, I cannot fail to appreciate for myself the features of your situation that appear to call out for outraged responses. I defend this claim against an
objection that stems from the apparent inarticulacy of some of our own emotions. The fact that we are not always sure what exactly we are angry, sad, or excited about might seem to suggest that our emotions can be unintelligible to us, but I explain why this is not the case. Then, in the final sections of Chapter Three, I turn to a defense of the claim that empathy is not just one means among others by which we secure humane understanding. I consider two challenges to this claim. The first denies that perspective taking is necessary for appreciating the intelligibility of others’ emotional outlooks, on the ground that this intelligibility can be secured through the mastery of folk-psychological narratives. The second denies that emotional engagement is necessary for appreciating the intelligibility of others’ emotional outlooks, on the grounds that emotion does no work to further this appreciation that could not be done by cold reason instead. In response to the first objection, I argue that empathy is in fact necessary for our mastery of narratives, and in response to the second, I contend that unemotional perspective taking provides us only distorted access to the intelligibility of others’ emotional perceptions.

With a sense of the epistemic import of empathy in hand, I return in the fourth chapter to the question of whether we can vindicate Shelley’s claim that empathy is “the great instrument of moral good.” Empathy has recently come in for some heavy criticism. It has been described as not only morally unnecessary, but positively “morally corrosive.” Paul Bloom and Jesse Prinz, the two most prominent critics of empathy, argue that empathy is good for much less than we might have thought. They contend that contrary to what we might have expected, empathy is not necessary or sufficient for our knowledge or understanding of moral properties, nor is it necessary or sufficient for moral motivation. The motivational and epistemic goods typically attached to empathy can all be

---

secured through other means. And, what is more, we should prefer those other means to empathy. Empathy is biased, narrow, and bad for empathizers’ psychological health.

Bloom and Prinz are correct in some respects. Empathy is biased, and we do not always need to empathize in order to discover the moral truth or become inspired to do the right thing. However, in focusing solely on empathy’s contributions to our knowledge of the good and to our motivation to act on that knowledge, they end up missing out on an important dimension of empathy’s significance. If we pay attention to the perspective of those on the receiving end of empathy, we will see that people have a powerful need to be empathized with, not (just) because of the possible downstream consequences for the empathizer’s desire to help, but because empathy relieves a painful form of loneliness. Empathy can be a way of caring, rather than just a means of getting us to care. I draw upon E.M. Forster’s *A Room with a View* and on Smith’s poignant remarks about the pain of being misunderstood in order to more precisely identify the sense of isolation for which empathy is the only cure. Empathy is valuable to us as a way of being in fellowship with other human beings, and (I further argue) it is especially valuable within the context of friendship. In fact, I suggest that the presence of empathy may be critical to distinguishing friendship from other relations of admiration or respect.

Empathy matters morally in part because people value being humanely understood for its own sake, but people do not always value being humanely understood in the same way. I explore several different sources of variation in the need for humane understanding, including concern for mental privacy and worries about the connection between empathy and pity. I also address the question of whether there are some perspectives that we ought not to empathize with, no matter how much their bearers want to be humanely understood. I suggest that we should resist the idea that empathy
is only virtuous if it is concerned with perspectives that are themselves either virtuous or, at worst, morally neutral. At a minimum, we have some moral reason to empathize with perspectives that are bad in a venial sense. Doing right by others may sometimes require those who would be virtuous to compromise the purity of their own minds, to think unlovely thoughts and entertain low desires.

After exploring empathy’s status as a non-instrumental moral good, I turn back to the two domains Prinz and Bloom focus on, moral epistemology and moral motivation. Empathy actually has more to offer us in both of these domains than Prinz and Bloom acknowledge. On the epistemic front, empathy has a meaningful role to play in the development of both our reflective beliefs about moral properties and our emotional sensitivity to those same properties. Empathy can allow us to discover new moral distinctions through imaginative emotional experience, and it can also help to shape and sharpen our moral emotional perception. On the motivational front, empathy with others makes it very difficult to intend to treat them in certain immoral ways. We cannot both empathize with someone and see them as unfeeling, as unthinking, or as lacking a normative perspective on the world. Consequently, empathizing with someone is at least not compatible with treating them as a mere object. Typically, it will also not be compatible with treating the target of empathy as a mere passive observer of the world, rather than as an agent capable of deliberation.

My description of empathy’s nature and significance will necessarily be incomplete. My hope, though, is that it will at least help to show that there are interesting questions to ask about empathy beyond the ones that predominate in the current literature. Sustained reflection on how empathy matters to us can help to expand our conception of what others need from us, and even drive us to rethink some of our assumptions about what is involved in being good.
CHAPTER ONE

A Portrait of Empathy

1. Introduction

My aims for this chapter are simple. I will begin by developing a more robust initial picture of the phenomenon Shelley briefly describes, which I am calling empathy. I intend this picture to capture the commonsense understanding of the phenomenon, the one largely shared by we who have experience of empathizing and of being empathized with. I will also highlight the points where consensus about empathy’s nature tends to break down. The end result will not be a sharp, clean, and complete description. Our sense of empathy is hazy in patches, and my characterization of it in this chapter reflects that. Still, what I have to say here about empathy will serve to distinguish it from other phenomena that have travelled under the same name, and I will point out the differences between my use of “empathy” and others that regularly appear in the philosophical and psychological literature.

The second task of the chapter is to introduce a popular conception of empathy’s significance that I will call the bridging picture. According to this picture, empathy offers a means of escape from the solipsism and egoism that would otherwise be our natural state, and that it does so by taking something of our way of relating to ourselves and what is ours, and transferring it to our relations with other people and what is theirs. Many early modern and contemporary philosophers (along with, perhaps, a handful of pre-Enlightenment thinkers) embrace this same idea: empathy matters because it leads us to knowledge of others’ minds and altruistic concern for their well-being. In the chapters that follow, I will be concerned with whether the bridging picture can be developed in a coherent and realistic way, one that distorts neither the nature of empathy nor the nature of our
other-oriented concern and understanding. To that end, I conclude this chapter by introducing a set of desiderata for any account of empathy’s nature and significance.

2. Empathy’s essential features

Shelley’s phenomenon, as he describes it, has the following features: it is a mental activity that (1) implicates vivid imagination (2) involves getting at others’ experiences “from the inside” and (3) somehow involves the emotional life of the empathizer. I will address each of these features in turn.

2.1. Imagination

Philosophical consensus about the features of imagination is relatively thin. Nevertheless, we can say some things about what imagination is that will help to make it clearer what is and is not included in our conception of empathy, without entering too deeply into controversial territory. First, imagining is active, in the sense that it is something we do, even though it is not always purposefully undertaken. The degree of deliberateness involved in imagination in general, and empathetic imagining in particular, is widely variable. Sometimes a dogged commitment to the project of imagining being in another’s position is necessary for getting empathy off the ground (think here of the kind of effort involved in trying to put oneself in the position of someone whose opinions and manners immediately strike you as rotten). Conversely, it sometimes also happens that we just find ourselves imagining how things look from another person’s point of view, without really setting out to do so. A simple gesture or glance is often enough to set our imaginations to work.

Shelley remarks that the greatly good person’s imagination is “intensive” and “comprehensive.” Presumably, less morally accomplished people can empathize too, even if their imagination is less powerful or inclusive. Not just anything that falls under the broad umbrella of imaginative activity
will be sufficient to satisfy the first criterion, though. We can, at least, contrast the kind of imagination that we (and Shelley) think is centrally involved in empathizing, what I will call *vivid* imagination, with mere supposition.¹ Talk of vivacity (famously associated with Hume’s model of mental association) frequently recurs in discussions of empathy.² But what does it mean for imagination to be vivid, in contrast to suppositional? When we suppose, we merely consider what would follow from the truth of either a counterfactual or a hypothesis. We entertain a proposition or set of propositions. “Picturing” how things would be is no part of supposition. By contrast, vivid imagining typically implicates various sensory modalities in an immersive mental creation or recreation of the target experience or phenomenon.

The easiest way to pick out the contrast is perhaps by example: the difference between merely supposing that I am on the 8:15 bus to Harvard Yard and vividly imagining being on the 8:15 is the difference between simply assigning a hypothetical positive truth value to the claim that I am on the bus, on the one hand, and picturing getting on the bus, grasping a cold rail, and hearing the brakes squeal as we draw up to the Holyoke Gate, on the other. In addition to having this kind of sensory dimension, vivid imagining might also bring desiderative and emotional states into play. I might imaginatively immerse myself in relief that I will make it to class on time, and in self-satisfied pleasure at my unusual timeliness. Notice that this is not the same as supposing *that* I am relieved or self-satisfied. Acts of supposition can concern emotional, desiderative, and sensory states (we can suppose that we are terribly sad, for example), but the imaginative indulgence in such states is unique

¹ “Vivid” imagination is sometimes called “enactive imagination.” See e.g. Goldman (2006): 47. The differences between supposition and vivid imagination are discussed at length in, among other places, Currie and Ravenscroft (2002) and Moran (2001): 75-106.

²See e.g. Nancy Sherman’s claim that in empathizing, “we see the particulars of others’ experiences with the vividness of being in their shoes” (1998): 117, and Ian Ravenscroft’s claim that in empathizing with a rock climber he “vividly experienced what it was like to be him” (1998): 171.
to vivid imagination. We will be concerned with what exactly that indulgence amounts to later on. As I will discuss in more detail below, in my description of empathy’s emotional dimension, there is a great deal of puzzlement as to whether vividly imaginatively encountering something as (say) terrifying involves really being terrified, or being “lightly” terrified, or simulating terror, or something else entirely. At this point, it is enough to have the basic (and, I think, intuitive) distinction between supposition and vivid imagination in view, and to say that the degrees of intensity and comprehensiveness by which we might measure acts of vivid imagination are not a matter of how many propositions we entertain, nor how focused on them we are, but instead a matter of how extensively we muster and deploy our various sensory, desiderative, and emotional sensitivities in representing the target of our imaginative efforts.

Typically, we conceive of imagining as involving creative work. Imagining being in a garden, for instance, involves conjuring up and bringing together mental representations of blossoming flowers, the twittering of birds, the feel of soil beneath one’s feet, and so on. Complex feats of imagination, like the ones involved in empathy, may involve a number of different mental operations: not only the combination of representations, but also subtraction and abstraction. In what follows, I will in general retain the conception of empathetic imagination as generative: it involves the creation of a new, complex representations. However, I want to acknowledge that the process of vividly calling to mind an image, whether or not that image is creatively stitched together, could also be considered imaginative in a broader sense.

---

3 The idea of vividly imagining as involving “lightly” feeling is due to Pugmire (2005): 185.

4 This broader sense corresponds to the notion of imagination common to many philosophers of the early modern period (though not, interestingly, Hume). Descartes, for instance, holds that “whatever we conceive with an image is an idea of the imagination” (in his letter “To Mersenne, July 1641”; see Descartes (1991): 186). Vivid recollection belongs to the Cartesian imagination just as much as flights of fancy do.
A final initial point about the nature of imagination: it may be the case that our folk concept of imagination allows for the possibility of imagining that which we take to be true. After all, we do say to each other things like, “can you imagine how beautiful that portrait must be?” thereby prompting each other to visualize what we take to be the actual beauty of an existing but absent portrait. And when we speak of “imaginativeness,” what we have in mind does not necessarily have anything to do with thinking about the non-actual: “imaginativeness” most centrally concerns the ability to uncover unexpected links, or to see things in an interesting, novel light.

Still, one standard kind of imaginative activity (call it counterfactual imagining) involves entertaining a possibility that the imaginer knows is not actualized. If, in the course of imagining, the imaginer loses her grip on that knowledge, her activity ceases to be what it was. Counterfactual imagination cedes its place to delusion. As for what counts as losing one’s grip on the non-actuality of what we are imagining: we might lose our grip when we are no longer willing to assent to the proposition “I am entertaining a non-actual possibility.” But this willingness is probably not the unique, definitive marker of the difference between counterfactual imagination and delusion. In some cases, we may maintain this particular willingness even as we behave in ways more consistent with accepting the counterfactual as actual.

Shelley falls just short of making an explicit claim that the relevant form of imagination for the phenomenon he’s celebrating is counterfactual imagination. However, it is not hard to see that that

---

5 There is a complication here. It might be argued that in such cases we must be imagining actually seeing the portrait, which of course we are not. I will not pursue this question.

6 A point made by Moran (1994b): 86.
is what he has in mind in his remarks. He is talking about taking others’ places, and presumably he does not mean that the greatly good man literally occupies many others’ positions. Henceforth, I use “imagination” to mean “counterfactual imagination.”

The claim that empathy involves vivid imagination already allows us distinguish it from several other phenomena that have sometimes shared its name. For one thing, it gives us some grounds to say that empathy in my sense is not the same as the simple feeling of distress at seeing another’s suffering which has at times been called “empathy” or “empathic distress.” The psychologists who use the term “empathic distress” to mean a feeling of distress at witnessing suffering generally accept that no imagination need be involved in the generation of empathic distress. The idea that empathy just is a feeling of distress at witnessing suffering obviously leaves open the possibility that there are non-imaginative routes to empathy, and that is a possibility I mean to foreclose.

Empathy also cannot be identical to the broader class of cases of an emotional response that, in Nancy Eisenberg and Janet Strayer’s formulation, “stems from another’s emotional state or condition and is congruent with the other’s emotional state or situation,” where “congruent with the others’ emotional state or situation” means, roughly, of sufficiently similar character to the other’s emotional state, or roughly appropriate to the other’s situation. Versions of this latter definition, some with an added “awareness condition,” are in wide use in both philosophy and psychology. For instance, Sober and Wilson define empathy in this way: “S empathizes with O’s experience of

---

emotion $E$ if and only if $O$ feels $E$, $S$ believes that $O$ feels $E$, and this causes $S$ to feel $E$ for $O$."

Nancy Snow takes a strikingly similar tack: “$S$ empathizes with $O$’s experience of emotion $E$ iff (a) $O$ feels $E$, (b) $S$ feels $E$ because $O$ feels $E$, (c) $S$ knows or understands that $O$ feels $E$."

These definitions will count among empathetic experiences some cases that my description of empathy as necessarily imaginative excludes, since congruent feelings may arise from a recognition of another’s emotional state that is not imaginative. Perhaps I know that you are annoyed, just because you have told me so, and I subsequently become annoyed at your annoyance. It seems plausible to say that we are in congruent emotional states and that mine has stemmed from yours. My emotion would therefore count as empathetic according to (at least) Snow’s formulation. However, there is no obvious role for imagination (especially of a vivid, counterfactual sort) to play in this story. I can be irritated by your emotional state without vividly picturing it to myself.

Imagination’s role also serves to distinguish empathy as I conceive of it from another phenomenon sometimes described as empathetic, one which was the special concern of late 19\textsuperscript{th} and early 20\textsuperscript{th} century phenomenologists such as Max Scheler, Edith Stein, and Edmund Husserl. These philosophers used “	extit{Einfühlung}” (empathy) to denote “a special kind of perception of the psychical states as they are manifest in the bodily expression” (Ingarden) or “a kind of act of perceiving sui generis…the experience of a foreign consciousness in general” (Stein).\textsuperscript{12} While the phenomenologists disagreed about the precise nature of this phenomenon, they all associated it with the perception or

\begin{itemize}
\item \textsuperscript{10} Sober and Wilson (1998): 234.
\item \textsuperscript{11} Snow (2000): 68. See the first of Batson’s eight conceptions of empathy in Batson (2009) and Hoffman (2001): 29 for other very similar definitions.
\end{itemize}
quasi-perception of another person’s mentality and mental states “in” the others’ gestures and expressions, the kind of perception or quasi-perception involved in (as Scheler puts it) being acquainted with “another person’s joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, with his entreaty in his outstretched hands, with his love in his look of affection, with his rage in the gnashing of his teeth.”¹³ I take the kind of experience the phenomenologists were interested in not to be a particular characterization of the mental activity Shelley had in mind, but rather a distinct phenomenon. In fact, the phenomenologists criticized previous philosophers’ efforts to invoke something like Shelley’s phenomenon in order to explain our acquaintance with other minds, and pointed to their empathy as a more realistic alternative. The phenomenologists’ conception of Einfühlung does invite the question: if it is true that we can perceive others’ mentality, their emotions, and even their intentions by looking at and listening to them, what is their left for imaginative empathy to teach us about others’ inner worlds? This is a question well worth asking, and it is one that I will aim to provide an answer to. For now, though, I simply want to note the difference between my conception of empathy and the phenomenologists’ conception(s).

2.2. Getting at others’ experiences from the inside

Having said a bit about imagination, let me now move on to the second feature I include in my conception of empathy: empathy involves getting at others’ experiences “from the inside.” Shelley refers to “putting oneself in the place of another.” The same metaphors of relocation come up again and again in descriptions of empathy: the empathizer “stands in the other person’s shoes,” or “trades places,” or “see things from the other’s position.”¹⁴ To draw out a bit more what it is,


¹⁴ The “shoes” locution, in particular, is everywhere. See e.g. Snow (2000): 17; Batson (2011): 34; Boler (1997): 257.
exactly, that we have in mind when we use these locutions, and to get a bit more precise about what we don’t have in mind, I will expand on what is meant by “others,” what is meant by “experiences,” and what is meant by “getting at from the inside.”

First, which others? Shelley implies that the creatures it is morally important to empathize with are members of “the species.” The reference to “the species” invites two questions: whether we can empathize with non-human creatures, and whether it is morally important to do so. It is hard to know exactly what to say about our ability to imaginatively inhabit the perspective of non-humans. Some think that we cannot really imagine what it is like to be a bat. Others are quite sure that we can imagine being anything that enjoys consciousness of some kind, and only definitively draw the line at rocks. Some of what I have to say later on should help us to think about whether we can empathize with non-human creatures, and also whether we ought to. Here, I will simply note the total lack of philosophical consensus on this point.

We can also ask whether empathy can extend to all members of “the species.” There is one individual whose status as a potential object of empathy is particularly tricky, namely oneself. If empathy does involve imagination, then it seems like one will not be able to empathize with oneself in any straightforward or typical sense: I can’t imagine being in my current position, because I am, as it were, already there. There are two kinds of special cases to consider, though.


First, there is the matter of past or future people that it seems in some sense right to call myself. Is the four year-old Olivia Bailey in principle an appropriate target of my empathy? Or the eighty year-old Olivia? I think it puts little strain on the ordinary conception of empathy (that is to say, Shelley’s phenomenon) to think that I can empathize with the distant future Olivia (if not the actual Olivia, then the imagined one). The case of the past Olivia is more complicated. I can remember what it was like to have broken my arm on the playground in 1989. Does remembering something interfere with the possibility of imagining it, thus foreclosing the possibility of empathy focused on that experience? The relation of memory to empathy is a vexed question, in large part because the question about whether episodic memory is itself a form of imagination is so hotly contested. What the case of episodic memory and the case of empathy undoubtedly have in common is that they both involve vivid picturing, and more specifically a kind of looking out at the world from a perspective that is not currently one’s own. If empathetic imagining necessarily involves the kind of creative combination of mental images I mentioned above, then one who embraces a conception of memory as itself generative in this sort of way should be inclined to think that far from crowding each other out, episodic memory (from the first-person perspective) and empathy with one’s past self will actually amount to just the same thing. One who holds that episodic memory is not generative, by contrast, will resist this identification, and less willing to accept the possibility that empathy could operate in the case of remembered episodes. I will not adjudicate between these possible positions here, but I will return to the matter of how one’s own episodic memory may contribute to empathy with others later in the dissertation.

17 For some recent important contributions to this debate, see Michaelian (2016), Debus (2014), Bernecker (2009).

18 So long as she accepts the characterization of empathetic imagination as itself generative in nature.

19 See pages 207-208.
Let me now mention the second kind of special case. Self-help books are full of advice to “empathize with yourself.” Consider, for instance, these recommendations: “[R]eal maturity comes with empathizing fully with yourself—first and foremost”; “When you hear that critical voice…you can empathize with yourself about how painful it is to always feel criticized, scared, less than others, hopeless.” The thought these passages seem to express, that their addressees both could and should empathize with themselves, here and now, will make more sense once I give the associated titles. The first quote is from *A Way out of Madness: Dealing with your family after you’ve been diagnosed with a psychiatric disorder*, and the second is from *Coping with Trauma-Related Dissociation: Skills training for patients and therapists*. Both of these guides address cases in which people are severely alienated from attitudes that nevertheless remain in some sense their own. Psychiatric trauma and disorder might create enough distance between parts of the self that one part can empathize with another, without performing that task so efficiently that it to longer makes sense to speak of parts of a single self at all. For this reason, and because of the other kind of case I’ve discussed above, the claim that empathy is always about the experiences of others, and not oneself, is too simple. However, in the cases where it seems possible to empathize with oneself, the target of empathy is in some sense encountered as other, as not totally identical to the empathizer.

Turning from the matter of “others” to that of “experience”: we often say that empathy is concerned with getting at “what it’s like.” “What it’s like” talk in philosophy is often concerned with narrow targets, specifically instances of sensory experience: what it’s like to see a certain shade of blue, or what it’s like to taste pineapple. Interestingly, though, that we seldom speak of empathy with others’ sensory experiences of redness or pineapple flavor. Why not? Well, we might be tempted by the response that empathy is generally directed at much broader and more general swathes of

---

experience, such as what it’s like to be a black man in America, or what it’s like to be in exile. But while it is true that empathy is often concerned with what it is like to lead a certain kind of life, it can also be directed at much narrower targets. Another person’s experience of a headache, for instance, seems like a perfectly normal target for empathy. What’s the difference between a headache and an experience of redness, though, that would account for the one being a typical subject of empathy, the other not?

Here, I believe, Adam Smith can help us out. He wrote: “‘Pity’ and ‘compassion’ are labels for our fellow-feeling with the sorrow of others. ‘Sympathy’ [NB: Smith is talking about the same thing I am calling empathy] though its meaning may originally have been the same, can now fairly properly be used to denote our fellow-feeling with any passion whatever” (TMS 10).21 By “passion,” Smith means feelings of joy, anger, resentment, delight, terror: in short, feelings which present their objects as either good and to be sought after and praised, or bad and to be avoided and condemned (or sometimes, in the case of complex emotional responses, a mixture of both). The idea that empathy is inevitably concerned with or directed at attitudes that represent their objects as broadly good or bad is perhaps too restricted. We might think that we can empathize with another’s experience of something as strange, for instance, and the experience of something as strange is not easy to fit neatly into either of the aforementioned categories. The key thought that I take from Smith, though, is that empathy is typically concerned with experiences of things or situations as evaluatively charged.

I will not attempt to draw a bright line between “merely” sensory experiences and experiences that are evaluatively charged. Still, experiences of objects as red or acidic are obviously not experiences of value in the way that experiences of delightfulfulness and hideousness are. To return, for example, to

21 All references to Smith are to the 1982 Liberty Press edition of The Theory of Moral Sentiments (abbreviated as TMS).
the case of the experience of a headache: when we empathize with someone’s migraine headache, we might imagine seeing a pulsing visual aura, feeling nauseous, and experiencing an intense pressure behind our ears and eyes. But if we were somehow able to vividly imagine these sensory experiences without appreciating the awfulness of the headache, it would be at least somewhat odd to say that we were empathizing with the migraineur. I believe this observation to be well supported by our own experience of empathizing and being empathized with, even though it does not standardly figure in contemporary philosophical and psychological characterizations of empathy. In Chapter Three, I will argue that the observation is, in fact, critically important for the proper analysis of empathy’s epistemic significance (I will also further defend it there).

I’ve said something about “others” and about “experience.” Rounding out the second major feature of empathy is the specification that it involves getting at others’ experience “from the inside.” I can encounter aspects of your experience not from the inside. This can, of course, happen through testimony, when you tell me what you are thinking about or how you feel. I can also encounter elements of your experience more directly, but not from the inside. When I see your crumpled face and hear your moans, and register them as a grimace and moans, respectively (rather than just as a set of noises and a particular geometrical configuration of flesh) I seem to encounter your pain directly. I see it and hear it. We can draw the distinction between encountering your experience from the inside and from the outside by thinking about an example of an imaginative encounter that is not, intuitively, empathetic. When we read about the woman burned alive for witchcraft in Paraguay in 2014, we can imagine what the scene must have looked like from a bird’s eye view: how

22 For an important exception, see D’Arms (1999), as well as (of course) Smith (1853).

23 As mentioned above, a number of phenomenologists defended the position that others’ mental states are directly perceptible (and, of course, the idea that we directly perceive mentality also figures in Wittgenstein’s *Philosophical Investigations*). For more recent defenses of versions of this claim, see e.g. Gallagher and Varga (2014), Dretske (1973), Cassam (2007).
the victim might have screamed, the expression on her face, and so on. This imagination could be intense and richly detailed, and it might elicit a powerful emotional reaction from us. In these senses, it looks empathetic. But it still won’t be a case of empathetically relating to the victim because it involves imaginatively approaching the scene from a vantage point that is not the victim’s own.

It is not entirely clear how we should understand the relation between the distinction I’ve just drawn and one that has particularly concerned psychologists, that between imagining what one would think and feel in the other’s position, on the one hand, and imagining the other’s thoughts and feelings, on the other. The motivational significance of this supposed difference has been a major focus of empathy researchers, most prominently Daniel Batson. According to Batson, his studies of subjects’ responses to being asked to put themselves in the shoes of a sufferer (imagine-self) versus being asked to imagine the other’s feelings (imagine-other) reveal that “an imagine-self perspective appear[s] to produce a mix of self-oriented distress feelings (tense, upset, etc.) and other-oriented empathic feelings, whereas an imagine-other perspective produce[s] relatively pure empathic feelings.”

24 In the study from which Batson drew this conclusion, subjects tasked with adopting the latter perspective were asked to “imagine how the other participant likely feels while waiting to hear which task he or she will be assigned, believing the tasks are to be assigned by chance,” and to “Imagine also how the other participant will likely feel when told which task he or she is to do.”

25 The problem, from a philosophical point of view, is that it is not clear whether the subjects who received and tried to follow the latter instruction imagined being the other person (a from-the-inside form of imagining), or whether they instead imagined the other’s experience from an outside perspective (or, indeed, whether they did something else altogether).

---


The psychologists’ distinction between “imagine-self” and “imagine-other” perspectives may or may not be the same as one that several philosophers have made between imagining being myself in your situation, and imagining being “in your situation as you.” It is not clear to me that there is a critical distinction to made at the point at which imagining being me tips over into imagining being you. It is true that there is a difference between imagining being myself transported with character, history, and physical features intact into Napoleon’s literal location at the battle of Waterloo, at the one extreme, and imaginatively taking on as many aspects of Napoleon’s physical, historical, and psychological circumstances as possible, in the manner of a method actor, at the other extreme. But these extremes exist on a scale with many intermediate grades. If I imagine being a 18th century Frenchman on a battlefield, one with Napoleon’s battle ambitions but also my own terror of violence, am I imagining being Napoleon, or me in Napoleon’s position? The question doesn’t seem to have an obvious answer, nor can we easily say why anything of moral significance should hang on what answer we give to it.

That is not to say, though, that there are not special worries about our imaginative assumption of certain features. Some circumstances are relatively easy to imagine. Many mundane shifts will fall into this category, of course, such as having a different hair color, or having gone to a different school. But even very complex and significant differences in circumstance don’t prevent us from vividly imagining. We can imagine, for instance, having a wildly different career, or growing up in a different age. Our imaginings may not be very accurate, or course, but a person with a normal

---


27 Goldie (2011) suggests we should reserve “empathy” for a case in which I transport nothing of myself into the imagined perspective, which does seem to have the effect of rendering empathy tremendously rare.
imagination will not be stumped as to how to go about imagining these things. One the other hand, there seem to be real puzzles about the limits of imagination along other dimensions. Can I really empathetically imagine being more or less intelligent than I am? Or will imagination in this case only extend to the supposition that I am so? How about imagining being more sentimental than I am, or less irritable? It seems that there is something about at least some psychological features of a person that, intuitively if rather mysteriously, places them beyond empathetic imagination. And, intriguingly, the features that are “beyond” in this sense aren’t obviously all and only the ones that are critical to our personal identity. We might plausibly think I would still be myself if I were a bit less intellectually acute, but empathetically imagining being less intelligent may nevertheless not be a genuine possibility. The question of how these limits on our imaginative abilities matter morally and epistemically will be very important in the final two chapters of this dissertation.

A final note about the claim that empathy involves getting at others’ experiences from the inside: in her influential *Caring: A Feminine Approach to Ethics and Moral Education*, Nel Noddings describes a phenomenon that seems to share this characteristic, but which she distinguishes from “empathy’ as it is usually defined” on the grounds that the phenomenon she has in her sights “does not involve projection into the other, nor does it imply a full understanding of the other.” She says of this phenomenon, which she calls “engrossment,” that it involves “receiving the other” such that “I become a duality.” I do not “put myself into the other’s shoes”; rather, “I see and feel with the other. …The seeing and feeling are mine, but only temporarily mine, as on loan to me.”

Do Shelley and Noddings have the same phenomenon in view? It is rather hard to say. Noddings’ talk of “receiving and sharing” another’s attitude, and her explicit denial that engrossment involves

---

asking ourselves questions like “What would I feel if I were wet to the ribs?” (the example here being engrossment with an unhappy baby in a bath) might suggest that although engrossment is supposed to involve getting at experience from the inside, it does not involve imagining, and so is not empathy in my sense. However, it is hard to conceive of how seeing partly through another’s eyes could be non-delusional unless it is imaginative, and while Noddings admits that we can “lose ourselves” in engrossment, she does not claim that we always do so; engrossment is not in itself a state of delusion. Unlike emotional contagion or the phenomenologists’ *Einfühlung*, engrossment as Noddings describes it may be imaginative and certainly involves getting at other people’s experiences from the inside. In these respects, at least, it looks like a close enough fit with our rough initial characterization of empathy that we should not assume she is describing some other phenomenon. Rather, we should consider Noddings’s description of engrossment to be a candidate characterization of how, exactly, empathy works.

2.3. The role of emotion

We can now turn to the third feature I find in Shelley’s description, namely empathy’s implication of the emotional life of the empathizer. I will be relatively brief with this element of our initial portrait not because this feature is more self-explanatory than the others, but because intuitive consensus runs out quickly on this point. Shelley claims that for a man to be greatly good, “the pains and pleasures of his species must become his own.” His phrasing does not make it clear whether another’s pains and pleasures “become one’s own” only if one actually experiences (1) those very same (numerically identical) pains and pleasures or (2) pains and pleasures that are tokens of the same type as those felt by the other or (3) pain or pleasure as the result of some grasp of the other’s

---


pains and pleasures, or if it is rather enough that one 4) “has” the other’s pains and pleasures in some other sense that does not actually involve experiencing them, but nevertheless makes them one’s own and/or involves one’s affective or emotional capacities. The use Shelley makes of “pains and pleasures” is also ambiguous in that it is not clear whether “pains and pleasures” refers to the feelings themselves, or rather to the things that elicit them (in the sense that your loss is your pain, or your triumph your pleasure).

The unclarity of Shelley’s remark is not accidental or idiosyncratic. It is commonly thought that empathy must involve the emotional capacities of the empathizer in some way. In imagining what the other’s experiences are like, it is generally agreed, we somehow exploit our own network of tendencies to feel and/or react emotionally. To develop a picture of how emotion is involved in empathy beyond this basic point of consensus, one needs to confront a pair of observations. They both seem obviously correct, and yet philosophers have found it difficult to fully accommodate them both in a single picture of empathy’s emotional nature.

The first observation is that empathy does seem to involve actual feeling. Empathy is not a cool, intellectual enterprise: people commonly speak of feeling the pain or frustration of the one with whom they empathize. The thought that empathy involves actually feeling (either as part of the imaginative process or otherwise) is not just prevalent amongst non-philosophers, either. It also shows up all over the philosophical literature in all manner of different guises. So, for example, we find Jesse Prinz, a thoroughgoing skeptic about empathy’s significance, writing: “empathy is a matter of feeling an emotion that we take another person to have.”31 Lawrence Blum describes the feeling

---

31 Prinz (2011): 215. This claim, incidentally, fits most easily with the second possibility I mentioned above (pains and pleasures that are tokens of the same type as those felt by the other).
involved in empathy in a different way, but one that again leaves no doubt that genuine feeling is
involved: “empathy involves a feeling for another, that is, a feeling whose object is another’s state of
mind or situation.” It is hard to even know how to begin to categorize Schopenhauer’s remarks on
empathetic feeling, but again, there is no doubt about the reality of the emotional experience. He
writes of our empathy with suffering: “it is precisely in his person, not in ours, that we feel the
suffering. We suffer with him, thus in him; we feel his pain and do not imagine that it is ours.”
Schopenhauer seems here to be suggesting that in feeling empathetically, we feel pains, but feel them
as not our own. We might be inclined to think that what he means by this is that we feel the other’s
pain “from the outside,” in something like the way you can “feel” someone else’s fever by touching
their forehead, except that he also says that we suffer with the other. This is just a sampling, of
course, but the unity within diversity on exhibit here, a common commitment to empathy’s
involving actually feeling, but a disagreement about how to characterize that feeling, more or less
reflects the state of a good portion of the literature.

The challenge for those who would defend a view of empathy as involving real feeling or emotion is
to square that view with the observation that the putative feelings involved in empathy do not seem
to bear the connections to action and behavior typical of non-empathetic feeling. Emotions are
typically linked to specific patterns of behavior and action. When we are disgusted, we attempt to
avoid the thing we find repulsive, to get it out of our sight. When we are angry, we tend to lash out
or otherwise act in a hostile manner. These ties to behavior are strong enough that if a person
claimed to be terribly frightened of, say, a nearby dog, but exhibited none of the behaviors typical of

32 Blum (2011): 172. See also Maibom (2014): 4: “to say that S feels E for O is to say that the E S experiences is
empathic.”

fright, we might be inclined to disbelieve them, or at the very least conclude that they could not be 
all that scared. Now, if like Blum, one maintains simply that the feeling involved in empathy is directed 
at another’s “state of mind of situation,” then this connection between emotion and action poses no 
problem for the claim that empathy involves genuine feeling. After all, joy directed at another’s (say) 
happiness could retain the usual connections between emotion and behavior without any oddness 
thereby arising: if I am joyful about your happiness, then I might well be motivated to announce my 
feelings, invite others to celebrate with me, do a little dance of jubilation, and so forth. However, 
most accounts of emotions’ involvement in empathy include more constraints on what counts as 
empathetic emotion than does Blum’s (and, I think, they are right to do so). Blum illustrates his 
account of emotions’ role in empathy with cases where the empathizer’s and the target’s emotions 
are is some sense matching, such as disappointment directed at disappointment, but his account is 
open enough to allow that even something like sadness at your excitement (perhaps because I think 
your excitement is foolish) could count as empathetic emotional engagement. This stretches the 
limits of our conception of empathy. We tend to think that the empathizer’s emotion is not (or not 
just) about another’s emotion, but rather corresponds to that emotion. And this is where the difficulty 
with emotions’ connections to action and behavior really emerges.

If empathy requires that I feel your suffering, your fear, your rage, such that I the empathizer do in 
some sense suffer, or fear, or become enraged, and if genuine emotions are typically connected with 
behavior along the lines I’ve just mentioned, then how can we explain the fact that one who 
empathizes with another’s fear does not herself cower, seek safety, or indeed exhibit any of the 
behaviors typical of a frightened person? Empathizing with another’s fear doesn’t typically involve 
this kind of behavior, and the same goes for the other emotions, too. We don’t expect empathizers 
to, as it were, act on or act out the emotions that the target of their empathy feels. We wouldn’t say
of a person who claims to empathize with another’s anger, but who does not, say, rage aloud or seek revenge, that she has empathized poorly or inadequately.

This puzzle about the typical behavior of empathizing individuals closely parallels a familiar problem from the philosophy of fiction. We seem to have emotional reactions to fiction, and yet those emotional reactions don’t typically issue in behaviors typical of the given emotion types (the classic example: we don’t rush to rush to console Anna Karenina, despite our sorrow at her plight). And a possible solution to the puzzle in the case of empathy is one borrowed from the discussion of this close cognate problem: one might characterize what the empathizer has as “quasi-feelings” or “pretend feelings,” in line with Kendall Walton’s description of emotions directed at fictional characters or events as “quasi-emotions.” These are on Walton’s view not genuine emotions, but rather “constellations of sensations or other phenomenological experiences characteristic of real emotions” that nevertheless lack the element of belief that makes a real emotion what it is. The absence of the element of belief, on this proposal, means that one’s (pseudo) emotional involvement with an imagined scenario is strictly quarantined. The simulation might involve our emotional capacities, but it will be contained such that the empathizer does not actually have feelings in imagining those experiences. This quarantine also means that when one is empathizing properly, the impact of pseudo-feeling or quasi-emotion is not allowed to spill over into the realm of

---

34 For some key contributions to the literature on this and related questions about emotional responses to fictions, see Walton (1990), Carroll (1990), Moran (1994b), and Gendler (2008).


37 One way we might want to put the point is that when we empathetically imagine, our emotional capacities are at work “off-line.” The term “off-line” originated with Robert Gordon (see Gordon (1986)) but he used it to signal the way in which simulated thought is prohibited from influencing the simulator’s action, rather than their emotion.
behavior.

In Chapter Three, I will propose that there is no need to describe the empathizer’s emotions as “pseudo-” or “quasi-” anything. But I will hold off on that argument for now, and note instead that the question of how emotion is implicated in empathy is further complicated by the recent trend, in taxonomizing forms of empathy, of maintaining that one form of empathy does involve feelings, but then positing a second kind of empathy that does not involve feeling at all. In his 1992 APA Presidential Address, a paper that helped set the agenda for much of the contemporary work on empathy, Alvin Goldman suggests that empathy has two “manifestations,” one “affective” and one “interpretive.”³⁸ Philosophers do commonly acknowledge that “[e]mpathy, as ordinarily understood, is affective empathy.”³⁹ Still, it is now quite typical to begin one’s analysis of empathy with a cleaving of “cognitive” empathy, empathy that amounts to something like “knowing another person’s internal state, including his or her thoughts and feelings” from “affective” empathy, typically defined in terms like these: “an emotional state triggered by another’s emotional state or situation, in which one feels what the other feels.”⁴⁰ Some philosophers have opted just for the less bold line that empathy has affective and cognitive dimensions.⁴¹ Even those who opt to distinguish between dimensions rather than kinds of empathy frequently follow the general trend by moving to isolate empathy’s cognitive and affective sides, though. There is a general sense that they can profitably be studied independently. and it is rare to encounter opposition to this approach. When Nel Noddings writes in the 2013 edition of her book *Caring* that “It is almost certainly a mistake to separate off

---


⁴¹ See e.g. Simmons (2014), Decety and Jackson (2006).
cognitive and affective dimensions of empathy,” she condemns what has become a standard philosophical move.\(^{42}\)

In addition to questions about the exact nature of the feeling involved in empathetic experience, then, we also may ask whether there is a purely cognitive counterpart to Shelley’s phenomenon, one that involves imaginatively but unfeelingly adopting the other’s perspective. And if there is not, or if there is, but its nature is not well-described by the taxonomizing efforts mentioned above, then it may be profitable to further ask: what prompted the move to distinguish between cognitive and affective empathy in the first place? Is it thought that our actual experience of imaginatively engaging with others’ inner lives naturally invites this distinction? Has the distinction been introduced in order to solve some problem? Or is it just a practical shift to allow for a division of labor, one that has seemed permissible in light of much more general beliefs about the separability of the cognitive from the affective? These are big questions, ones we will make some progress toward answering in the chapters to follow.

2.4. Emotional contagion

I have now fleshed out my description of each of the empathy’s three key features, in a way that I hope reflects our commonsense conception of Shelley’s phenomenon. One remaining point that deserves a moment of consideration is the relation between empathy as I conceive it and a phenomenon that is variously said to be the inevitable precursor of empathy, a proper part of empathy, one route among others to empathy, or even just an alternative to empathy.\(^{43}\) Michael

\(^{42}\) Noddings (2013): 231. For another notable exception, see Deigh (1995).

\(^{43}\) For the claim that emotional contagion is an alternative to empathy, see Coplan (2011). For the claim that it is a primitive form of empathy, see Darwall (1998). For the claim that it is empathy’s precursor, see Maibom (2014).
Slote, who has long been concerned to restore empathy to a place at the center of moral theory and moral life, explains that the target of his research is “associative, receptive, or emotional empathy, which occurs when we are invaded, so to speak, by the feelings or attitudes of another person.”

This kind of talk of invasion frequently shows up in characterizations of a phenomenon often named “emotional contagion.” What is emotional contagion, and what is its relation to empathy? Can empathy involve the “invasion” Slote has in mind, or does the invasion have no part to play in the activity of inhibiting another person’s perspective? Since I will be dealing in the following chapter with a view of “sympathy” that looks very much like the thing Slote has in mind, it is important to achieve some measure of clarity on these points.

Whether or not they include emotional contagion under a broad umbrella of empathetic response, characterizations of emotional contagion generally place an emphasis either on (1) the distinctiveness of emotional contagion’s causal mechanism or (2) the special character of the subject’s own understanding of their emotional state. Let’s begin with a standard characterization of the first sort. Justin D’Arms describes emotional contagion as an “involuntary ‘catching’ of another’s emotion, induced somehow by the model’s expressive behavior.” In line with most contemporary philosophers and psychologists who take an interest in the matter, he suggests that the principle mechanism by which this kind of contagion occurs is mimicry, especially that of facial expressions of emotion. A substantial body of scientific literature indicates that facial mimicry is a hard-wired, automatic, and primitive human behavior, and that furthermore, this mimicry tends to trigger in the mimicking individual emotions congruent with those expressed by the other’s facial movements.

---

44 Slote (2015): 14


46 For a brief, helpful survey of this literature, see Hatfield et al (2006).
D’Arms’ characterization of emotional contagion stresses its non-deliberative, involuntary nature. We are doing the mimicking, to be sure, but it is not something we embark upon reflectively or otherwise undertake to do.\textsuperscript{47}

In contrast to D’Arms’ account, other accounts of emotional contagion stress the importance of the subjects’ apprehension of their own experience. Amy Coplan, for instance, underlines that “with emotional contagion, the emotions are not experienced imaginatively or in relation to another; we experience them as our own,” and Douglas Hollan points out that “for many researchers” what makes empathy different from emotional contagion is the empathizer maintains a “clear cognitive and experiential boundary between the two [the empathizer and the object of empathy], such that the empathizer can always distinguish between her own thoughts and feelings and those of the other.”\textsuperscript{48}

The connection between these two ways of characterizing emotional contagion is obvious enough. If emotional contagion is a primitive process that produces emotion in us automatically, without any effort on our parts, then it is unsurprising that we who experience emotional contagion would typically be unaware of the etiology of our emotion, in which case a knowledge of the emotion’s etiology could not color our experience of it. In a typical case of contagion, I “catch” the blues of another person in the room without being aware of it, and I recognize my newfound gloom without yet having an accurate sense of how I came by it. I’m even liable to engage in a bit of confabulation, and conclude that my blue feelings stem from and are concerned with the limpness of my lunch salad and the greyness of the day. That said, it certainly seems that one could catch an emotion from

\textsuperscript{47} For another description of emotional contagion very similar to D’Arms’, see e.g. Nichols (2001).

another, along the lines of the causal mechanism D’Arms describes, without remaining ignorant of or mistaken about the source of the emotion. If I realize that what I am feeling now is a “caught” emotion, that might well change my experience of it. Merely knowing the causal history of my emotion is not guaranteed to change anything about its role in my mental ecosystem. But if I can see my caught sadness as an echo of another’s sadness, I can experience it “in relation to another,” as a representation of the emotion of another. In that case, my emotion will have a contagious provenance, it will be the product of an emotional “invasion,” but it will have come to mean something quite different to me than it did initially.

An emotion that is contagiously acquired, and that is not experienced “in relation to another” in the way just described, will not count as an empathetic emotion on my view. An emotion with a contagious provenance that is re-interpreted as a reflection of another’s emotional experience, on the other hand, could indeed count as empathetic. In saying this, I don’t intend to make a special exception for these kinds of experiences, one that allows them to count as empathetic while being excused from the usual rules about (1) being produced through imaginative activity and (2) being an instance of getting at others’ experiences from the inside. This is because the reinterpretation the contagious emotion undergoes effectively transforms it by placing it within a new context. In the kind of process I am envisioning, when I successfully reinterpret my contagiously acquired emotion, I am no longer just sad. Rather, the sadness is framed by a conception of my feeling as reflecting what it’s like for her. In this context, my sadness will naturally be shaped by my imaginative representations of her situation. My experience contains, as it were, a persistent through-line of sadness, but my sadness may take on a very different meaning for me once I appreciate its nature and history.

I’ve been working to develop a picture of Shelley’s “great instrument” that highlights the points of
lay and philosophical consensus about the nature of empathy, while also making it clear where the most serious uncertainties lie. Perhaps unsurprisingly, some of these uncertainties tie into major, longstanding debates in philosophy of mind and epistemology: the difference between memory and imagination, the accessibility of others’ consciousness, the relation between the cognitive and the affective. I of course cannot pretend I will be able to resolve these debates satisfactorily, so I cannot promise to replace all the relevant hazy spots in my picture with perfectly sharp and distinct details as I go. Still, now that we have in hand an idea of empathy’s basic elements and a sense of ways in those elements have been further described in the philosophical and psychological literature, we are well enough placed to look at how we think empathy matters.

3. How we think empathy matters: the bridging picture

[F]or each of us the universe is deeply and sharply and completely divided into for example in my case, me, on one side, and everything else, on the other. This for each of us exhaustively defines the whole universe.

David Foster Wallace, The Broom of the System49

Suppose that our world were drained of empathy. If we were to cease imaginatively entering into others’ perspectives, feeling our way into worldviews not our own, would our relations with others, or indeed with ourselves, be impaired in any way worth caring about? And if so, would we lose anything that could not be supplied by other means?

Philosophers of mind, moral psychologists, sociologists, and cognitive scientists disagree widely about how empathy matters. Nevertheless, among those who affirm empathy’s significance, there is one strikingly broad consensus about what would be missing from an un-empathetic world.

Empathy, it is widely agreed, is the only way (or, in less bold formulations, a critically important way) we have of overcoming a gap between ourselves and other people. Amy Coplan characterizes empathy as the means of rescue from epistemic isolation. When I empathize, she suggests, my previously unique relation to my own mental “island” is extended to apply to your territory as well.\(^{50}\) Schopenhauer, similarly, describes empathy as the way of overcoming the “wide chasm” egoism posits between self and other.\(^{51}\) Other authors opt for the less dramatic “gap” or “divide” in order to describe the space between persons to be overcome or “bridged” by empathy.\(^{52}\) The characterization of the chasm varies depending upon the interests of the author. Epistemologists tend to describe a gap in knowledge or understanding, whereas moral psychologists talk about a gap of concern. Still, the geographic motif remains the same. Without empathy, it is suggested, we would each be stranded on our own private landmass, isolated from all of our fellows.

This “cartographic” vision of how empathy matters derives its appeal in large part from two basic observations about human relationships. The first is that our own thoughts and cares occupy very different positions in our mental economy than do others’ attitudes and needs. The full story about these positions will be complex, but we can sum up the idea in a very rough, general way like this: our self-knowledge and self-concern seem both more secure and more intensive or extensive than our knowledge of and concern for others.\(^{53}\) The thought that our epistemic and motivational position with regard to ourselves is different because it is privileged has for many provided support for

\(^{50}\) Coplan (2011): 18.

\(^{51}\) Schopenhauer (2010): 263.


\(^{53}\) Throughout this chapter, I am using “knowledge” in a loose sense, without distinguishing between propositional knowledge, acquaintance, understanding, and the various other epistemic goods in the immediate vicinity.
the further thought that our self-knowledge and self-concern is more primitive than our other-oriented self-knowledge and self-concern. The second basic observation is that despite these differences, we often succeed in becoming intimately acquainted with others’ inner lives, and we are sometimes powerfully moved by the plight of others. This second observation, when taken in conjunction with the first, creates a puzzle that empathy seems perfectly fitted to solve, to wit: how is it that we are sometimes able to get outside of ourselves, and overcome the differences we observe between self-and other-oriented concern and knowledge?

Let me now flesh out those two observations a little. I will begin by briefly surveying some often-noted differences between self-knowledge and knowledge of others, and also between self-concern and concern for others. How, exactly, one tells the story about empathy’s traversal of the self-other gap will depend in part upon what one identifies as the differences between our self-orientation and our other-orientation. My intention here is not to determine which of these differences are important, or which of them empathy might actually be capable of overcoming. I only want to bring the bridging picture and its attractions more clearly into view.

Let us first consider the epistemic front. Ryle was poking fun when he wrote that “The mind is its own place and in his inner life each of us lives the life of a ghostly Robinson Crusoe,” but there is some intuitive appeal to the idea that we are, in our inner lives, able to encounter our own mental goings-on in ways that outsiders cannot.\textsuperscript{54} That picture has been developed in various ways. In some cases, the insider’s epistemic advantage has been principally characterized as special knowledge concerning the very existence of our own mental life. In \textit{On the Basis of Morality}, for instance, Schopenhauer claims that we are naturally bound to know that our mental lives are real, while the

\textsuperscript{54} Ryle (1949): 3.
fact of others’ mindedness remains a shakier proposition.\textsuperscript{55} Other accounts describe our knowledge of our own selves as unique in that we have a singular grip on the contents of our beliefs, thoughts, and desires. It has been claimed that we have exceptional inferential access to the goings-on of our inner lives, that we can know our mental states through introspection, that we can know our minds in a special way because we alone make them up, and so forth.\textsuperscript{56}

It would be overstating matters to say that there is philosophical consensus that we are in a special epistemic position with regard to our own inner life. Some philosophers are willing to affirm that our self-knowledge is not importantly different from our knowledge of others.\textsuperscript{57} However, such “no-difference” views at least face an uphill battle when it comes to reconciliation with our commonsense convictions. If the thought that my grip on my inner life is different from my grip upon others’ is an illusory one, it is at least a very difficult one to shake. Moreover, the popular conception of the difference has it that my self-knowledge is not just distinct but also privileged in at least some respects. It is sometimes the case that others know us better, for some sense of “know,” than we know ourselves. A psychologist might have a better understanding of the true source of my resentment than I do, and a marketing specialist will quite likely be better able to predict my reactions to a supermarket display than I will. But we can grant them their expertise and still insist that we have an important inside track of some kind (or, perhaps, inside tracks plural: there might be multiple, mutually irreducible ways in which our epistemic relation to our own mentality is special).

\textsuperscript{55} Schopenhauer (2010).

\textsuperscript{56} For two different kinds of inferentialist view, see Byrne (2005) and Carruthers (2011). For views that focus on our special ability to “make up” our minds (sometimes called agentialist views), see e.g. Burge (1996) and Moran (2001). For a helpful summary guide to the theoretical space see Gertler (2015).

\textsuperscript{57} See e.g. Ryle (1949), Stich (1983).
There is a lot of disagreement about the exact nature of these tracks or tracks. However, two linked features that philosophers attempting to analyze self-knowledge have frequently acknowledged and tried to account for are the apparent comparative immediacy and security of much of our knowledge of our own minds. If I deliberate about whether \( p \), and conclude that \( p \), it does not seem that in order to know my own mind on the subject of \( p \) I now will have to do something further, namely investigate whether there is a belief that \( p \) knocking around in my consciousness. I certainly won’t have consult my own external manifestations of my belief (my gestures, my behavior) in order to divine that I believe \( p \), nor will I have to ask myself: do I believe \( p \)? My warrant for the conclusion that I believe \( p \) appears to be that, as far as I can see, \( p \).

My knowledge of my own belief in such cases seems to come for free, and in such a way that there is really no room for uncertainty about my own mental state to take root in a gap between my believing \( p \) and my knowing that I believe \( p \). The apparent lack of an uncertainty-generating gap accounts for the security of much my knowledge of my own mind. While it may be possible for me to be wrong about some of my beliefs (those buried deeply in my subconscious, for instance), my knowledge of the content of my conscious thought does not depend upon the comparatively precarious inferences I might draw from my behavior, or my testimony, or generalizations about what people tend to think in such-and-such circumstances. The story looks quite different, though, when another person concludes that \( p \). Another’s conscious conclusion that \( p \) can be hidden from me. I can be very wrong about what they think. And that’s no mere hypothetical possibility. The

\[58\] This description won’t be accepted by everyone, of course. Someone who holds that all mindreading (self-directed and other-directed) is accomplished via the same mechanism of sensory-data interpretation (e.g. Carruthers (2011)), or who holds that we are aware of our own mental states via “scanning” processes (e.g. Armstrong (1993)) will not accept this claim: my concluding \( p \) means I need not now seek or process evidence (beyond the evidence that \( p \)) of my belief that \( p \) in order to know my own mind. But even someone who thinks that our self-knowledge is inferential can embrace my characterization; Byrne’s “self-verifying” inferential scheme (\( p \to I \) believe that \( p \)), for example, affirms that latter claim and does not allow for an uncertainty-generating gap (see e.g. Byrne (2005)).
insecurity of our knowledge of other’s minds is the source of much everyday confusion, as well as the fodder for comedy and tragedy. And a similar asymmetry pertains, not accidentally, to immediacy: it is not the case that once the other person has concluded that \( p \), the warrant for my belief that she believes \( p \) is necessarily in place. She might believe that \( p \) without my thereby securing any justification for the conclusion that she believes that \( p \).

Those apparent differences in the manner of knowing may be paired with differences concerning the volume and extent of one’s knowledge. While we sometimes insist that others “know us better than we know ourselves,” it seems plausible to say that I have (or have access to) much more knowledge of my mind (at least my conscious mind) than does anyone else. Every single conscious thought I have may readily be converted into a bit of self-knowledge. As for extent, one might think that there is exactly one person whom I naturally or originally know “what it’s like” to be. I know how blue looks to me, and what agony or joy or jealousy feels like to me, and so on. In short, I know the quality of my own experience of the world. Once I’ve experienced some sensation or appetite, I know what it is like for me. This knowledge of the quality of experience is practically a given when it comes to my own encounters with the world. Conversely, it has been argued that it is mysterious how we manage to even conceive of the experiences of others, let alone know them.\(^{59}\) It seems we must be able to, and yet, once we take seriously the thought that all our direct knowledge of experience is knowledge of our own experience, worries about how we can even conceptualize others’ experience begin to creep in.

So, to sum up the apparent differences: We seem to know our own minds more, more immediately, and more securely than we know other minds. We also seem to know, practically inevitably, what our

\(^{59}\) This is the “conceptual problem” most famously associated with Wittgenstein (1953). See also Nagel (1974).
experiences are like, whereas our knowledge of other’s experiences remains at the most imperfect and spotty, and perhaps (if one resolves the conceptual problem by bullet-biting) impossible. Now let me turn to the differences between our concern for ourselves and our concern for others.

There are two things we might mean by “concern.” First, there is the interest we take in something. When we are concerned about or with something, it occupies our attention. We worry about it, scheme about it, dwell on it. In this sense, concern can be malicious or benevolent. I will primarily be focused on a narrower sense of concern. When we are concerned in this narrower sense, the object of concern occupies our attention and we are motivated to act for the sake of it.

When it comes to the difference between self-concern and our other-directed concern, philosophers often emphasize a difference in magnitude rather than a difference in manner. Adam Smith memorably claims that our feeble general love of humanity is not enough to prevent us from preferring the destruction of the whole world to that of our little finger (TMS 136). And a clear echo of that thought is preserved in Schopenhauer’s remark that “[E]goism, both in an animal and in a human being, is linked in the most precise way with his innermost core and essence, and indeed is properly identical with it…. [E]goism is colossal. It towers over the world. For if the choice were given to any individual between his own destruction and that of the world, I do not need to say where it would land in the great majority.” 60 Smith and Schopenhauer are thinking of the contrast between ourselves, on the one hand, and all the nameless teeming masses that populate the earth, on the other. And when we are comparing self-concern with concern for strangers, there is in general a striking difference in degree. The story does not look so dramatic when it comes to friends and family, of course, but while many people are very strongly motivated to aid their friends, it is almost

certainly true that they are much less concerned in that first, purely attention-directed sense with their friends than they are with themselves. After all, almost all people have to make choices and move themselves through the world, and that demands attention.

It can be hard to tease apart differences in magnitude from differences in kind. But it seems like we don’t just happen to be more concerned about ourselves: our self-concern also seems immediate and inevitable in ways that our concern for others does not. As for inevitability: Certainly almost all of us recognize, in some sense, that the suffering of the world merit our aid. We affirm that their oppression, their starvation, their violent deaths are a bad thing, and we affirm that we really ought to do something. But those affirmations are often quite disconnected from a sustained orientation of concern. We are, it turns out, entirely capable of living with the thought that others are in agony, without rushing to relieve them. Conversely, the prospect of registering one’s own agony without experiencing an intense aversive reaction is an utterly bizarre one. When it comes to our own occurrent agony, indifference seems to be possible only in cases of extreme psychological disturbance, while relative indifference to the ongoing agony of others, especially distant others, is perfectly banal. It is also quite unusual to be unmotivated by the prospect of our future agony, although that state is perhaps not as rare and baffling as indifference to our own actual ongoing agony. It is also worth mentioning that even the other-oriented concern we do rouse ourselves to is frequently fleeting, whereas self-concern hardly ever ceases to do its work.

As for immediacy: when we move to alleviate our own agony, it seems like we are hardly ever motivated by the thought that this is what we ought to do. Still less are we typically motivated by reasoning that we should act to relieve our pain because we are ourselves a being of intrinsic moral worth, or a human being, or anything of that sort. Our motivation to respond to the pains and
pleasures we are actually experiencing is usually more reflexive than reflective. In fact, when it comes to self-concern, we tend to think that one who needs to avail herself of the reasoning that she is a human, and that all humans deserve concern (or some such line of thought) in order to motivate herself to relieve her own suffering must be severely depressed or else strangely alienated from herself. In parallel with the case of self-knowledge, there seems to be no gap between registering one’s agony and being motivated to put an end to it. By contrast, it is at least not very unusual for our motivation to aid another to depend upon our having reasoned our way from a recognition of suffering to the conclusion that helping is now the thing to do, via some mediating principle. In sum: our self-concern generally seems stronger, more durable, closer to inevitable, and more immediate than our other directed concern.

Together, these observations about self-knowledge and self-concern naturally suggest a picture according to which primary forms of robust self-concern and self-knowledge are more primitive than knowledge of and concern for others. I will call the thought that self-knowledge is more primitive than other-oriented knowledge “epistemic egocentric primacy” and the thought that self-concern is more primitive than other-oriented concern “motivational egocentric primacy” (and the combination of these two thoughts simply “egocentric primacy”).

The claim that our self-concern and self-knowledge are “more primitive” could be understood in a number of ways. Versions of egocentric primacy that show up in the philosophical and psychological empathy literature tend to feature one or both of the following senses of “primacy”: (1) developmental priority: children possess at least some forms of self-concern and self-knowledge before they possess the corresponding forms of other-directed concern and knowledge, and their eventual other-directed concern and care is best understood as emerging at least partly from self-
concern and self-knowledge, (2) general explanatory priority: the other-directed concern and knowledge of even mature adults is best understood as partly emerging from their self-concern and self-knowledge. In both cases, the thought is that we must move from what is practically inevitably given (self-concern, self-knowledge) to that which is more tenuous and uncertain.61

The claim that we start off in a kind of egocentric condition comes in more and less drastic forms. An extreme picture would have it that, pace Donne, every man is initially an island, entire of itself. Our primitive state is one of total solipsism and egoism. Less radical versions of this picture characterize our primitive knowledge and concern for others as comparatively circumscribed or weak, either in general, or along certain particular dimensions. The view that our primitive position is doubly (that is, motivationally and epistemologically) egocentric was a hallmark of much Enlightenment thought, endorsed to some degree by Locke, Berkeley, Hume, and Rousseau.62 In *Emile*, Rousseau even suggests that we are locked in this primitive position well into our teenage years: “At sixteen the adolescent knows what it is to suffer, for he himself has suffered; but he scarcely realizes that others suffer too…[he] knows no ills but his own…he [is] indifferent, like every child, to every one but himself.”63 Thanks to increasing philosophical specialization, the claim that

---

61 For claims of developmental priority, see e.g. Hoffman (2001) and Eisenberg and Strayer (1987). For claims of general explanatory priority, see e.g. Maibom (2007) and Slote (2015). Sometimes it is not clear which of these types of claims is being made: see e.g. Noddings (2013): 39, “[A] sense of my physical self, a knowledge of what gives me pain and pleasure, precedes my caring for others.”

62 Locke on self-care: “What is it that determines the will in regard to our actions?...[U]neasiness a man is at present under” (1843): II. xxi, 219. Locke on self/other knowledge asymmetry: “Man, though he has a great variety of thoughts, and such, from which others as well as himself might receive profit and delight, yet they are all within his own breast, invisible and hidden from others, nor can of themselves be made appear” (1843): III.i.3, 219. Berkeley on self-care: “Self-love being a principle of all others the most universal, and the most deeply engraven on our hearts, it is natural for us to regard things as they are fitted to augment or impair our own happiness” (1901): 104. Berkeley on self/other knowledge asymmetry: “[I]t is plain that we cannot know the existence of other spirits, otherwise than by their operations or the ideas by them excited in us…[T]he knowledge I have of other spirits is not immediate, as is the knowledge of my ideas” (1871): 231. Hume’s double egocentrism will be at issue in Chapter Two.

63 Rousseau (1921): 183.
our primitive state is *doubly egocentric* is less prominent today. The problem of our knowledge of others’ inner lives has become principally the province of epistemology and philosophy of mind, whereas the possibility of altruism is a concern for moral philosophers. Still, the thought that one or both of these egocentric conditions is a natural starting point for humans, one that we need to be drawn out of or “escape” from, lingers on. Nancy Sherman, for instance, closely echoes the enlightenment philosophers when she writes that “Our most basic abilities to understand others— to access their minds, to identify with their plights, to resonate with their joy” rests on the “transcendence of the self.”

On Sherman’s picture, the self is a given; it is where we start from. And she is not alone in this, although many contemporary philosophers will only be willing to endorse significantly weaker versions of egocentric primacy.

There are, of course, those who think that we are permanent castaways, incapable of ever really escaping the egocentric starting point. But here we come to the second basic observation about human relationships that helps to make the cartographic picture of empathy’s significance attractive: it certainly does seem that an escape from the island is possible, and indeed that it happens all the time. In the real world, it is rare to find a mature person with an outlook like that of Rousseau’s adolescent. Most of us, most of the time, have a pretty good grip on both the existence and the contents of others’ mental lives. We can be very intimately acquainted with the innermost thoughts of our fellows, and at least some of the time, we do seem to know a good deal about what others’ experiences are like. And we are also capable of fierce concern for other people. Even the most ordinary people, who hardly spare a thought for the needy and suffering in their everyday lives, will

---


65 Interestingly, the idea that we really are trapped within motivational egocentrism still has its philosophical defenders (see e.g. Mercer (2001) and LaFollette (1988)), whereas the corresponding epistemological thesis is rarely (never?) defended.
make tremendous sacrifices for others’ sakes. We abandon our dreams, give away our fortunes, and jump on grenades to ensure other people’s survival and happiness. That is to say: it seems like there at least some very reliable or even inevitable features of self-knowledge and self-concern that can apply (perhaps in an altered form) to their other-directed counterparts. The exact nature of those features is up for debate, of course, but the general observation is one that should have broad appeal.

The challenge for those who accept some version of the “islander” picture is to explain how we are able to get beyond our initial isolation. Empathy naturally recommends itself as the bridge or vessel that makes this transition possible. The idea that it might be the tool for this task is hinted at even in the initial characterization of empathy I’ve extracted from Shelley’s remarks. Recall: Shelley proposed that in empathizing, our usual relation to our own “place” is swapped for a relation to another’s place, and our relation to our pleasures and pains is extended to encompass to pleasures and pains of others, which become “our own.” Shelley casts empathy as taking something of our way of relating to ourselves and what is ours, and transferring it to our relations with other people and what is theirs.

Empathy looks like a bootstrapping mechanism, one that could exploit what is originally special about our self-directed knowledge and concern in order to make possible the insight and the benevolence that help to give our social lives their particular form. The hope is that the story we can tell about empathy’s bridging function will acknowledge and accommodate dramatic differences between our concern for and knowledge of others and our concern for and knowledge of ourselves, while also explaining how these differences can on occasion be overcome. The difficulty, though, is in working out the details of how empathy might do this essential work. We need the answers to
some important questions: how are the purported epistemic and motivational bridging functions related to one another? Can drawing on our own self-concern and self-knowledge really get us out of primitive egocentrism, rather than embedding us further in it? Will sharpening our sense of the nature of primitive egocentrism help to put meat on the bones of the bridging proposal, will it cause it to collapse, or something in between?

4. Desiderata for an account of empathy’s nature and significance

With the (still partially blurry) portrait of Shelley’s phenomenon now in hand, and with the bridging proposal concerning empathy’s significance in place, we can turn to the question of what we should look for in a satisfactory account of how empathy matters. First, a satisfactory account of empathy’s significance should be sensitive to the ways that we who empathize and are empathized with value it in practice. In treating such sensitivity as desirable, I am not closing off the possibility that many of us could be at least somewhat wrong about empathy’s good and bad qualities. However, if the significance an account ascribes to empathy dramatically fails to line up with our practical valuation of empathy, then a convincing error theory would be urgently required.

I think that philosophers’ inattention to the ways in which the recipients of empathy value it, in particular, has led to a distorted and incomplete picture of the goods empathy might supply, and the needs it might meet. At least in the domain of moral psychology, the philosophical focus has been on whether or not empathy is instrumental in securing goods whose value has already been well explored in other contexts, goods such as motivation to aid, or accurate prediction of behavior. Those goods are by no means inconsequential, but by taking people’s experiences of empathy’s value seriously (I will suggest) we will also uncover goods whose value it might be harder to pin
down outside of the context of empathetic experience. So, sensitivity to the ways we actually seem to value empathy is desirable not just because we want to “save the phenomena,” but also because it offers a means of correcting for a kind of narrowness in the current philosophical analyses of empathy’s value.

It would also be good for an account of empathy to reveal the significance of each of empathy’s characteristic features, as I’ve described them in the first half of this chapter. Empathy is imaginative and emotionally charged, and it involves getting at others’ experience from the inside. Shelley’s remarks capture the common sense that these features are all relevant to how empathy matters. However, many philosophical descriptions that present empathy as a morally and/or epistemically important phenomenon end up treating one or more of these features as insignificant. Some philosophers have concluded that it does not much (or not at all) matter to empathy’s hallmark epistemic and/or moral contributions that empathy involves perspective taking (Hutto and Gallagher (2008)), that it is imaginative (Walton (2014)), or that it involves the emotional life of the empathizer (Heal (2003)). It could be that there is nothing special about empathy per se, that it matters only insofar as it is a member of a larger class of phenomena (perspective-taking responses in general, say, or phenomena that involve feeling the same way other people do). In that case, were the world to be drained of empathy, it is possible that no goods would be entirely lost to us. I myself will argue that all of empathy’s features are necessary for the provision of empathy’s characteristic goods. Others who reject that stance should provide some account as to why we commonly but erroneously think that empathy, in all its particularity, is especially valuable.

\[66\] In this respect, we have a great deal to learn from anthropological research on empathy. Very closely observed accounts of people’s experience of empathy may be found in e.g. Throop (2010) and Hollan (2014).
Thirdly, an account of empathy should clarify the relation between empathy’s two purported dimensions of significance, epistemic and moral. We need to answer these questions: Is empathy both morally and epistemically significant? If so, is it morally significant in virtue of its epistemic significance, or vice versa? Or are there two completely separate stories to tell about the genesis of its epistemic and its moral significance? It is a striking and intriguing fact that empathy is said to be good in these two different senses, but many philosophers and psychologists do not attend to the possible relation between these goods, and among those who do, there is very little agreement about the nature of the relation. We ought to resolve that mystery, if we can.

Finally, it would be good for an account to identify the root causes of any significant distortions in thinking about empathy encountered in the literature. Uncovering where things went wrong is not just a matter of historical interest. If an alternative account of empathy offers an unrealistic picture of empathy’s significance, one that cannot be reconciled with our lived experience of empathy, that outcome will give us some reason to think that the assumptions or principles underlying the account are themselves worth (re)examining with a skeptical eye.

5. From the preliminary portrait to an enlightenment conception

In this chapter, I’ve tried to sketch a portrait of empathy that people who empathize and are empathized with will find familiar. I’ve also introduced the “bridging” conception of empathy’s significance, and done some work to demonstrate its attractions. Finally, I’ve identified the features that a satisfactory account of empathy should exhibit. This identification of features will guide the work to come. In the next chapter, I will turn to one enduring version of the bridging conception of empathy’s significance. It gets its first real articulation in Hume’s Treatise, and the treatment of empathy in that work will be my primary focus, but it is picked up by other sentimentalists, both
historical and contemporary. Close attention to the problems that affect the Humean picture will help us to identify the pitfalls a successful account of empathy’s significance must avoid, and will provide us with a better sense of the shape our account must take.
CHAPTER TWO
Humean Empathy: an Idea and its Afterlife

1. Introduction

My task in this chapter is to explore the earliest systematic attempt to explain how the tendency to make “the pains and pleasures of [our] species [our] own” matters. David Hume’s first attempt to describe the operations of “sympathy” appeared in print in 1739, in A Treatise of Human Nature. Hume sometimes uses the term “sympathy” to denote phenomena other than the one that Shelley had in his sights, but that fact has not prevented his views from continuing to exert an important influence on contemporary debates about the significance of empathy. His account’s capacity to explain recent experimental observations is frequently remarked upon, and some philosophers explicitly borrow Hume’s terminology in defining empathy.¹ More importantly, the ideas advanced in the Treatise about empathy’s epistemic and moral significance figure centrally in the recent revival of interest in empathy. Contemporary philosophers and psychologists may not have Hume explicitly in mind when they argue that, for instance, empathizers’ “experience [of] their own versions of [another’s] pain and disgust” renders benevolent concern for that pain and disgust “guaranteed.”² However, as we will see, the logic at work behind such claims often closely resembles the reasoning of the Treatise.

It is sometimes said that Hume is the father of one strain of research and thought about empathy, one that is concerned with the contagious transmission of sentiments, while his friend Adam Smith is the progenitor of a second strain, one that is instead concerned with the simulation of others’

mental states. That picture does not do full justice to the complexities of either philosopher’s views. Still, these two are undoubtedly the patron saints of philosophical and psychological work on the subject of empathy. I think it is important to begin with Hume, rather than Smith, because Hume’s more fully articulated philosophy of mind helps to bring into relief the problems that will emerge in pursuing what has seemed to many an attractive project, namely the demonstration that we can by means of empathy bootstrap ourselves up out of primitive egocentrism and solipsism. Hume’s account seems to end up ensnared in intractable difficulties, difficulties that are inherited by the contemporary theorists who have (unwittingly or not) pursued accounts of empathy’s significance that take after Hume’s. Once these problems are in view, I will turn to Smith’s account, and extract from it an alternative model of empathy’s importance (in Chapter Three).

I will begin by surveying the basic features of Humean mental architecture, so that we will be positioned to see how (and, to some extent, why) Hume accepts a version of the picture of egocentric primacy. As we have seen, this picture, versions of which continue to inform much of the contemporary work on empathy, generates mysteries to which empathy (or “sympathy”) is supposed to be the solution. Hume takes it that the process by which we arrive at a sympathetic end state furnishes crucial epistemic and motivational benefits. Through it, we come to know others’ emotional states and also develop a motivating interest (often benevolent, though sometimes malicious) in other people’s fates. Hume’s account of sympathy is complex. Therefore, I devote the second section of this chapter to a close consideration of his rather obscure and tangled remarks about how sympathy frees us from solipsism and egoism. As it turns out, there are two main explanatory strands that we can extract from his writing on the subject. I consider each in turn, and

---


4 A fact which even those who lament the inadequacy of their accounts readily concede. Scheler (1970): xlvii is a prime example.
highlight the respects in which each strand ends up being deeply unsatisfactory.

In drawing out these difficulties, I also indicate the ways in which they affect more recent analyses of empathy’s significance. After all, my ultimate purpose in examining Hume’s treatment of sympathy in the Treatise is not to show that Hume made mistakes. Rather, it is to uncover the problems inherent in an enduringly popular approach to analyzing empathy’s significance. Accounts of empathy that at least partially follow the Humean line end up running aground for different reasons. They end up making it look as though empathy simply contributes to the expansion of egoism, or as though empathy depends for its motivational efficacy on a kind of epistemic mistake. Or they avoid making empathy’s functions hinge upon epistemic error, but at the cost of making it look as though empathy is really an idle wheel, rather than a critically important bridge between oneself and others. At least some of these problems are known to the developers and defenders of such accounts, but as I will show, the ways offered of getting around them end up unacceptably distorting the phenomenon of empathy, or introducing significant explanatory lacunae.

The failure of Hume’s (and his descendants’) accounts to show how empathy can do the bootstrapping work we thought it might is not sufficient reason to give up the thought that empathy plays a critical role in bridging differences between self-oriented care and knowledge and other-oriented care and knowledge. But it is a reason to think again about exactly which differences empathy is in a position to bridge, and also to reconsider the relation between what empathy is supposed to do for our epistemic position and what it is supposed to do for our other-directed concern.
2. Egocentric primacy in the *Treatise*

In this section, I will I furnish evidence for Hume’s commitment to what I am calling the picture of egocentric primacy, according to which our knowledge of and concern for ourselves and our own inner lives is privileged, in virtue of being uniquely immediate and inevitable. To that end, I will review the relevant features of his model of the mind, including the nature of passions and the distinction between ideas and impressions.

Hume isn’t the first philosopher to speculate about our emotional responses to the perception of others’ suffering or joy. In the *Republic*, Plato notes that others’ emotions seem to call out for us to join in: “You know that we enjoy [hearing tragic heroes’ laments], give ourselves up to following it, sympathize with the hero (*sumpashontes*), take his suffering seriously” (*Republic* X, 605d). The invitation to go along with a feeling also memorably figures in Montaigne’s 1580 essay “On Cruelty,” in which the avowedly softhearted author confesses: “I am tenderly compassionate of others’ afflictions, and should readily cry for company, if, upon any occasion whatever, I could cry at all. Nothing tempts my tears but tears.”

It is easy to trace the influence of this essay in the work of Enlightenment philosophers; Montaigne’s description of a crowd’s emotional response to an execution, the witnesses’ reacting as though they felt themselves “in” the body of the hanged man, neatly aligns with Hume’s observation that the “rays of passions, sentiments and opinions may be often reverberated” between men (SBN 365). But neither Plato’s nor Montaigne’s discussions of “transferred emotion” (and what is to be done about it) have the programmatic character of Hume’s account. Hume elevates observations about our tendency to “mirror” others’ emotions to a central place in his philosophy of mind and moral psychology.

---

5 Montaigne (1910): 181.

6 All references to the *Treatise* are to the Selby-Bigge Nidditch edition (henceforth SBN).
Some Hume scholars have maintained that actually gives us two very different such accounts, one in the Treatise and another in the Enquiry. Here, I am going to focus exclusively on the considerably more detailed vision of “the principle of sympathy” articulated in Books II and III of the Treatise. It is quite possible that Hume’s decision not to recapitulate this vision in the later Enquiry was due to his own recognition of the problems I am going to highlight below. As we shall see, though, the flaws affecting Hume’s account of sympathy were not confined to the “dead-born” Treatise. They continue to recur in descriptions of empathy that take some of their cues from him.

2.1. Hume’s philosophy of mind: the basics

To get a grip on Hume’s conception of sympathy, we first need to revisit the basic tenets of Hume’s philosophy of mind. Hume opens the Treatise with the claim that all the perceptions of the mind are either impressions or ideas. Impressions are distinguished from ideas by their greater “force and liveliness,” and include “all our sensations, passions, and emotions, as they make their first appearance in the soul” (SBN 1). Ideas are inevitably copies of impressions, or (in the case of complex ideas) combinations of copies of impressions. Ideas generally resemble the impressions from which they derive in “every particular,” save for their “vivacity” or “force” (SBN 2). The former may on occasion approach the latter in vivacity, such that we mistake the one for the other, but Hume thinks that the two are generally quite easily distinguished. One might naturally regret that Hume didn’t offer a richer characterization of the difference between impressions and mere ideas: the words “vivacity” and “force” recur, but receive little further articulation. It may be helpful to think of ideas as sun-bleached versions of one’s original impressions, which retain the same form as the original but lack its attention-grabbing vibrancy.
It is by means of this phenomenological distinction that we separate ideas from impressions, but Hume also notes that ideas and impressions belong to different inner operations. Ideas are the province of thought, and are employed in our reasonings, whereas impressions belong to the realm of feeling. To complicate matters, ideas are further divided into ideas of memory and ideas of imagination. This division, too, is made solely according to variation in strength and clarity, rather than content. Ideas of memory, like the impressions of the senses, are (to use Hume’s language) “attended by belief,” and objects of belief are experienced quite differently than figments of imagination are: a believed idea feels more forceful, vivid, solid, firm, or steady. Conversely, ideas of imagination are not believed, and so have a different, weaker, “feeling to the mind” (SBN 269).

While believed ideas are relatively vivacious, however, it is important to note that an idea attended by belief is not itself an impression. I may infer, and believe, that there is a crocodile under the boat, but that does not mean I have an impression of the crocodile.

It is of critical importance to Hume that the vivacity of our perceptions is not static. The vivacity of one impression or idea may be transferred to a previously faint idea according to the links that make up various associative relations, including those of cause and effect and resemblance (the two that are particularly relevant to sympathy). The mind’s attention is inclined to move involuntarily along the links forged by these relations, such that the liveliness of one perception is conveyed to another like water flowing between pools along worn gulley-tracks. So, for example, given an impression of a portrait, the mind will move to the resembling idea of the person pictured therein, and the idea of that person will be enlivened. Associative mechanisms often just serve to bring different ideas to mind, but an infusion of liveliness can also vivify an idea already attended to, or even transform a mere idea into an impression. This possibility of this latter conversion will prove essential to the operation of sympathy.
2.2. The nature of the passions

So much for the most basic elements of the Humean mind. The mechanism of sympathy is particularly concerned with passions. Where do they fit into this scheme? Hume describes passions as secondary impressions, which is to say impressions that “proceed from” impressions of sensation. Impressions of sensation are a wide-ranging category, including tickles, scents, bodily pains, and bodily pleasures. But Hume claims it is only impressions of pleasure or pain that give rise to and sustain passion: “Tis easy to observe, that the passions, both direct and indirect, are founded on pain and pleasure…upon removal of the pain and pleasure there immediately follows a removal of love and hatred, pride and humility, desire and aversion, and most of our secondary or reflective impressions” (SBN 438).

The generation of a passion from impressions of sensation may proceed in one of two ways. So-called direct passions arise immediately from the experience of pain or pleasure, whereas indirect passions require the interjection of some further idea for their production. Typically, the requisite interjection is the association of the cause of our pleasure or pain with some person, either ourselves or somebody else. Love, hatred, envy, pity, pride, and humility are passions of the latter kind. It is easy to see why pride, for instance, counts as an indirect passion; to feel proud, I must associate some pleasurable thing with an idea of myself. In the case of indirect passions, Hume insists upon the distinction between cause and object: the inciting cause of a passion may not be identical to that thing upon which the passion focuses one’s attention, the object of the passion. The cause, for example, of pride might be one’s dog, but because the passion of pride “once raised, immediately turn[s] our attention to ourself,” the object of pride is oneself (SBN 278). Desire, aversion, grief, joy, hope, and fear, conversely, all fall into the category of direct passions, passions which arise as an
immediate response to the experience of pain and pleasure (SBN 277). In the case of direct passions Hume does not discuss the object/cause distinction, presumably because he takes it that the two will be largely identical: my fear is caused by the prospect of a painful bite, and directs my attention toward the very same thing.

For Hume, passions are actually all entirely “simple and uniform impressions” (SBN 277). They cannot encompass any perceptual or judgmental elements, since they are have no parts and are non-representational. Hume is clear on this point: “a passion is an original existence…and contains not any representative quality, which renders it a copy of any other existence or modification. When I am angry, I am actually possessst with the passion, and in that emotion have no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high” (SBN 415). The “very being and essence” of passions is constituted by “their sensations” (SBN 286). That said, passions do have objects, and because Hume allows passions to have objects, we can fairly speak of passions as being about things, without doing to much violence to Hume’s characterization of them.

2.3. Hume’s egocentric primacy

Hume’s characterization of passions as sensations entails that the physical manifestations of our emotions, our grimaces, chuckles, wrinkled brows and so forth, are not themselves part of passions, for those manifestations are of course not sensations at all. And that, in turn, helps to explain why Hume thinks we do not simply directly apprehend others’ passions, even when we hear their cries and see the tears streaming down their cheeks. Hume explicitly states that we cannot directly, non-inferentially know the identity of others’ mental states: “No passion of another discovers itself
immediately to the mind. We are only sensible of its causes or effects” (SBN 576). When it comes to our own passions, though, our knowledge is quite different. We have to do a little gap-filling on behalf of Hume, here, because he seems to think that the nature of this knowledge will be obvious to his readers once it is established that passions are sensations. Hume describes our acquaintance with our own sensations (and thus our own passions) as “constant” and “intimate” (SBN 339). Those labels suggest that we know our own feelings practically inevitably, and with the intimacy of direct experience rather than the remoteness of inference. Hume does admit that some passions of one’s own, the so-called “calm passions,” are, like others’ passions, “more known by their effects than by the immediate feeling or sensation” (SBN 417). But in making this exception he reveals what he takes to be the general rule: we feel our own (and only our own) passions, and when we feel them, we know them immediately. This picture of the asymmetry between the knowledge of our own and others’ passions places Hume squarely in the camp of those who endorse what I have called epistemic egocentric primacy, the view that we know our own mental states more, more immediately, and/or more securely than we know others’.

Hume’s also embraces the idea of motivational asymmetry when it comes to our concern for others and for ourselves. He explains that our motivation can generally be understood as deriving from the vivid perception of pain and pleasure: “There is implanted in the human mind a perception of pain and pleasure, as the chief spring and moving principle of all its actions” (SBN 118). We are immediately and forcefully moved by our impressions of pleasure and pain (which is to say, our feelings of pleasure and pain). And thanks to their great vivacity, our perceptions of our pains and pleasures dominate our desires: “nothing can be more real, or concern us more, than our own sentiments of pleasure and uneasiness” (SBN 469).

He repeats this claim in his discussion of our judgment of others’ motives: “This [the judgment of motives] we cannot do directly; and therefore fix our attention on actions, as on external signs” (SBN 477).
Hume does sometimes indicate that in addition to an original, unwavering desire for our own pleasure, we also have a non-derivative desire for the happiness of some others. He describes our love of our friends and desire for their happiness as “a natural impulse or instinct, which is perfectly unaccountable” (SBN 439). It seems that those to whom we are closely tied by the bonds of friendship or blood are loved naturally, and that Hume does not think that this love needs accounting for (SBN 352). We might think, then, that Hume goes in for a kind of half-hearted egocentric primacy, according to which our concern for ourselves and some others is primitive, whereas our concern for strangers is non-primitive and needs explaining. However, this interpretation is hard to square with the uncompromising boldness of the following passage, which is also notable for its reference to Hume’s views about mental privacy:

The sentiments of others can never affect us, but by becoming, in some measure, our own; in which case they operate on us, by opposing and increasing our passions, in the very same manner, as if they had been originally deriv’d from our own temper and disposition. While they remain concealed in the minds of others, they can never have any influence upon us: And even when they are known, if they went no farther than the imagination, or conception; that faculty is so accustom’d to objects of every different kind, that the mere idea, tho’ contrary to our sentiments and inclinations, would never be able to affect us. (SBN 593)

Here, Hume claims that if sympathy does not do its work, we will be indifferent to others’ sentiments (all others’ sentiments, it seems). And that strongly implies that the concern for our own passions that is so natural to us does not have an equally original and immediate other-directed counterpart. Can one both be indifferent to others’ sentiments and full of an “original” benevolent concern for others? Perhaps in a very limited sense; one might believe, in an abstract way, that others ought to be supported and protected, without caring about how they actually feel. But the tension between the two attitudes is nevertheless acute. And echoes of the claim that we need sympathy in order to arrive at a motivation to support others’ good, especially the good of strangers, are peppered throughout the *Treatise*. Hume remarks that our concern for “humans and sensible
creatures” “proceeds merely from sympathy,” and that “we have no such extensive concern for society but from sympathy” (SBN 481; SBN 579). So, I do not think we should hold back from saying that Hume goes in for a version of motivational egocentric primacy, despite the ambiguity about the status of concern for friends and family.

We have already seen why Hume holds that others’ passions are not immediately knowable, in the way that our own are: passions are sensations, and while the one who experiences a sensation is (generally) directly aware of it, others cannot see or hear or touch or feel that sensation. We can extract from the above passage a sense of the corresponding reasoning in the motivational case. Why does Hume claim that while the motivation to act for the sake of our own pleasure and pain is primitive, our motivation to act for the sake of others’ pleasures and pains is not? Because only very vivid perceptions of pleasure or pain (or, indeed, of anything else) are capable of motivating. Only a very vivid perception is able to, as it were, get itself taken seriously by a mind filled with an endless supply of ideas that cannot be acted upon. Because we feel our passions, they command the attention of our mind. And because we do so effortlessly and as a matter of course, our passions naturally and inevitably shape our motivations. Since we do not simply and directly feel others’ emotions, by contrast, Hume thinks we do not have the kind of vivid perception of them that could immediately direct our motivation. Hence the endorsement of motivational egocentric primacy, the view that our self-concern is inevitable and immediate in a way that our concern for others is not.

Hume’s picture of the mind generates a mystery. Hume clearly maintains that we are able to know others’ minds. And he is also very concerned to show that we care about others’ inner lives, and not in a way that is covertly egoistic. As a matter of fact, we do love other people, and that love is
invariably bound up with a concern to ensure other’s happiness.\textsuperscript{8} But given that we do not have immediate, vivid perceptions of others’ emotions, how can we possibly ever know them or be moved by them? How can we advance beyond the original state of being closed off to others, motivationally and epistemically, that the picture of egocentric primacy posits?

The language that Hume uses to present his theory of mind is somewhat unusual. His distinction between impressions and ideas, in particular, is not often echoed today. But it is not hard to get pulled into the basic problem he needs to solve. Many of the basic thoughts that lead him to his conception of the asymmetry between self-oriented care and knowledge, on the one hand, and other-oriented care and knowledge, on the other, are quite familiar, and could even be said to have the ring of common-sense. It is eminently natural to think that others don’t experience my sadness in the immediate and direct way that I do. And it similarly seems obvious that while my own misery persistently demands, as it were, that I do something about it, such that the possibility of experiencing pain without aversion seems almost unthinkable, I can contemplate another’s pain with relative indifference. Even if some moralists are inclined to insist that the latter attitude is aberrant or psychopathic, the fact remains that it is a possibility for us, and one that is often realized.

Egocentric primacy is not a view that expired with the Enlightenment. While increasing philosophical specialization seems to have pushed philosophers away from engaging with egocentric primacy as a dual problem, one that pertains both to knowledge and to care, both motivational and epistemic primacy are widely independently embraced to some extent. In fact, some contemporary philosophers are willing to cast the differences in more extreme terms than Hume did, especially when it comes to the motivational egocentric primacy. For example, Michael Slote, whose Humean

\textsuperscript{8} Invariably though, Hume is keen to emphasize, contingently. For Hume, love and desire for others’ goodwill are not connected through anything like conceptual necessity. See SBN 368.
analysis of empathy’s power I will return to later, argues that “if we don’t posit empathy, then the motivation behind or for altruism seems difficult or impossible to fathom,” on the grounds that the idea of motiveless malice is as unintelligible to us as wanting a saucer of mud, and “[i]f the idea of sheer raw malice makes no sense, why should the idea of sheer raw benevolence make any more sense?”

3. The two-phase operation of sympathy

Let me turn now to the Humean solution to this problem. Hume repeatedly identifies sympathy as the unique means by which we get “far out of ourselves” (SBN 579). His conception of sympathy does seem to shift somewhat over the course of the book. Hume’s best remembered lines about sympathy suggest a view of sympathy as a very mechanical communication of emotion: “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” (SBN 576). This description reveals the influence of Malebranche’s hydraulic conception of emotional transmission: “Even wonder…produces an expression upon our face that mechanically arouses the wonder of others, and that even acts on their brain in such a highly regulated way that the spirits it contains are forced into the muscles of their face to form an expression on it very much like our own.”

We do tend to associate Hume with the Malebranchian view, but if Hume were to stick with a view of sympathy according to which it operates via a simple, automatic reverberation, he would obviously not be entitled to cast sympathy as in itself sufficient for releasing us from primitive

---


egocentrism. For one thing, one might not even be conscious of the source of one’s reverberative feelings, their status as sympathetic. A “caught” feeling might easily be mistaken for an original one: I might feel uneasy because of your pained expression, but think that my feeling is simply a response to my own woes. For another, even if we add that one must be aware that one is sympathizing in order for sympathy to have the relevant bridging effects, sympathy conceived of in this way would not be able to furnish us with a detailed idea of other’s mental state. Hume doesn’t strictly think of emotions as intentional, but one important thing we might expect we could learn about other people’s inner lives is what kind of things their feelings direct their attentions toward. So, for instance, we think we can discover not just that the person next to us is angry, but that their anger has the slow wait-staff as an object, rather than, say, the war on terror. The transmission that happens when we automatically mimic another’s sad moue, and feel our upbeat mood draining away, is not what we might call very “high bandwidth”; we might come to know via this means that the other is sad, but it will not get us to an idea of the object of their sadness.

The contemporary empathy literature tends to overlook it, but the fact is that Hume doesn’t consistently describe sympathizing as a matter of mere contagion, akin to the phenomenon of infectious yawning. He also furnishes a more sophisticated account of sympathy as a process by means of which we vivify an idea that is of another’s experience. Strictly speaking, “sympathy” is for Hume a label attached to the second part of the two-phase process we are about to consider. However, these phases are naturally paired, and both involve bootstrapping up from something that is immediately available in the case of one’s relation to oneself in order to overcome relative ignorance of or indifference toward the other’s inner life. What is more, in conformity to my initial definition of empathy, both phases seem to involve the conception or experience of a passion as if from the inside.
Here is the most clear formulation of Hume's two-phase view, which I will refer to as the *focal passage*.

When any affection is infus'd by sympathy, it is at first known only by its effects, and by those external signs in the countenance and conversation, which convey an idea of it. This idea is presently converted into an impression, and acquires such a degree of force and vivacity, as to become the very passion itself, and produce an equal emotion, as any original affection. (SBN 317)

This explanation is of course highly telegraphic, but with the help of a few other remarks the proposal should become clearer. First, consider our acquisition of the idea of another’s passion. The focal passage emphasizes that we begin with a perception of the other’s gestures and exclamations, but elsewhere Hume allows that we can also start from a perception of “the causes of any emotion,” such as the perception of sharp surgical instruments, and the promise of painful surgery that they suggest (SBN 576). Once we have the causes or effects of emotion in view, the work of those associative connections comes into the story. Humans resemble each other very closely, both in the “fabric of the mind” and in the structure and compositions of our bodies. Crucially, we tend to move and speak in the same ways (SBN 318). Since our mind’s attention tends to flow along lines of resemblance, the idea or impression of another person’s gesture or words is likely to bring to mind the idea of our own corresponding gesture or word (your curled lip makes me think of my own).

To arrive at the idea of another person’s disgust, an additional associative mechanism is required, namely the association of cause and effect. Our own words and gestures are reliably connected to diverse inner states of our own, inner states with which we are immediately, non-inferentially acquainted. When we are disgusted, we tend to draw back and wrinkle our noses, and when we are delighted we tend to draw up the corners of our mouths. Therefore, we conceive of those inner states as the cause of those behaviors: the idea of my own curled lip is linked to the idea of my
disgust. And given the aforementioned resemblance between human beings, our minds are naturally carried to an idea of a similar cause at work behind the other’s behaviors. I infer that my companion is disgusted. He is experiencing something like the feeling of disgust with which I am familiar from my case. At this point, if I have not formed a mistaken associative link (as would be the case if, say, my companion’s gestures were just an act), Hume would say that I know my companion’s feelings. We can note, here, that this account fares better than the primitive infectious model in at least one way. Hume suggests that reflection on causes of passion may be part of our generation of an idea of another’s pain. Since the causes of pain are often identical to the things passions direct our attention toward, this model offers a way of thinking about how we can form ideas about the objects of others’ emotions.

Hume’s view of our knowledge other minds relies on an inference based on observed regular connection between causes or effects and emotions, but the relevant observed regularities are entirely restricted to one’s own case. Prior to sympathizing, the only idea of passion we have is the one that is the faint copy of our own impressions, our own feelings of anger, joy, despair, and so on. So, the only representation we can possibly conjure of another person’s passion is first-personal in character: we have no conception of pain “from the outside” that we could deploy. What exactly could the representation of the other’s passion produced by the first phase described in the focal passage look like, then? Hume does not say, but presumably it is a complex idea consisting of a faint copy of a first-person experience of a passion, combined with something like a “tag” that indicates “and this feeling is not in me, but instead over there, in her.”

One well-known problem with inferential approaches to knowledge of other minds like Hume’s is that they rely on a notoriously dubious logic. The fact that we outwardly resemble each other in
many ways does not seem like a very good reason to conclude that our passions (which are for Hume strictly inner sensations) are similarly correlated with external causes and effects. Hume says that we are furnished with knowledge via this string of associations, but we might doubt that the requisite epistemic justification can be secured through the associative links he posits. However, I want to pass over this relatively well-trodden problem in order to focus on the relation between the epistemic and motivational upshots of sympathy, so let me now turn to how sympathy is said to move us.

The idea of my companion’s state that the aforementioned associations secure is relatively lively. After all, the idea is one that we believe in, rather than one we are merely entertaining, and for Hume, belief in an idea amounts to having that idea with enhanced liveliness. But it is not yet a feeling. Hume does not think we actually need to be sharing the other’s sentiments, in the sense of actually experiencing something like the other’s emotion, in order to be aware of them: he later notes that others’ feelings can be “known” without “alone be[ing] able to affect us” (SBN 593). We have not yet reached the end of the sympathetic process, though. The second step in the process, as described in the second half of the focal passage, is the further enlivening of the idea of the other person’s regret to the level of an impression. This does result in our having an actual experience of a passion.

Hume argues that any idea associated with our own person is very lively: “[T]is evident, that the idea, or rather impression of ourselves is always intimately present with us…Whatever object, therefore, is related to ourselves must be conceived with a like vivacity of conception” (SBN 318). This means that the idea of my own disgust, which is brought to mind via the associative mechanisms just discussed, is very lively, and since the idea of my companion’s disgust is linked by resemblance to the idea of my own disgust, the liveliness of the idea of my own disgust is transferred
to my idea of the other’s disgust. Apparently, this transfer is so successful that my lively idea of the
other’s disgust will be “converted” into an impression (SBN 319). There is little room for doubt that
the sympathetically generated impression is itself a feeling: Hume says that through this mechanism
others’ passions become “our own” (SBN 593). My conception of the man’s pain, then, will
transform into a feeling, one that I myself experience. And this conversion is the lynchpin upon
which the birth of altruistic concern is supposed to turn.

Suppose I, the sympathizer, have experienced this conversion. Now I have a feeling, and feelings are
necessarily motivating. But how does this motivating feeling relate to care for others? Hume explains
in a passage that ranks among the most perplexing in the Treatise. Its full complexity can only be
brought out by quoting it at length:

When the present misery of another 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; possible, probable or certain. By means of this lively
notion I am interested in them; take part with them; and feel a sympathetic motion in my
breast, conformable to whatever I imagine in his…[w]e must consider, that benevolence is
an original pleasure arising from the pleasure of the person belov’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. In order, then, to make a passion run parallel with
benevolence, ’tis requisite we shou’d feel these double impressions, correspondent to those
of the person, whom we consider; nor is any one of them alone sufficient for that purpose.
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 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, as has been already observ’d; 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. (SBN
388)
There is absolutely no critical consensus concerning the interpretation of this passage, which has been described even by admirers of Hume as “rambling” and “not well written.” I think that its thorniness can be accounted for, though. Hume here mixes together two importantly different descriptions of how sympathy might yield benevolent concern. I suspect that he does so because both of the accounts are independently problematic. Unfortunately, as we will see, conjoining them in no way eliminates their flaws.

4. The joy of sympathizing

Let me now draw out the first such description I find in the above passage, which I will call the “joy of sympathizing” account. For Hume, sympathy does not invariably generate benevolence. Sometimes, it instead produces resentment. And what he has to say about the connection between sympathy and resentment will help to expose one way he is thinking about the sympathy/benevolence connection. He writes that when we sympathize with only one (painful) impression, “the sympathy is related to anger and to hatred, upon account of the uneasiness it conveys to us.” The original sympathetic emotion, the impression that the idea of the other’s pain turned into, must itself be an impression of pain. Hume clearly holds that the vivification of ideas is not a transformation that alters their content. How do we get from feeling sympathetic pain to other-directed resentment?

---


12 In fact, he regards this consistency of content as a feature that enables the conversion: “The component part of ideas and impressions are precisely alike. The manner and order of their appearance may be the same. The different degrees of their force and vivacity are, therefore, the only particulars, that distinguish them: And as this difference may be removed, in some measure, by a relation betwixt the impressions and ideas, it is no wonder an idea of a sentiment or passion, may by this means be inlivened as to become the very sentiment or passion” (SBN 319).
Hume suggests that the experience of sympathetic pain is unpleasant and aversive. He does not say why that is so, but it can’t be that it is unpleasant for us because and insofar as it is a reflection of another’s pain. He is trying to account for our hatred of others’ pain, after all, so he cannot appeal to it in his explanation of how this hatred arises. Presumably, we hate feeling sympathetic pain for the same reason that we hate feeling regular pain: the sensation of pain is intrinsically aversive. So far, our passions are directed only toward our own experience, and not toward any other’s. But we also know the ultimate cause of the unpleasant experience we are undergoing, namely the pain of the other. So, we are naturally inclined to resent the other as the cause of our discomfort. Conversely, we can assume, the joy of experiencing sympathetic pleasure, combined with the knowledge that the ultimate cause of our pleasure is the other’s pleasure, leads us to love the other as the cause of our joy. Love is for Hume not strictly identical to benevolence, but he does think that the two are closely linked. It is a contingent but universal fact that we want the good of those we love.

One obvious concern with this picture, as I have sketched it so far, is that it does not explain how sympathy with others’ pain and sorrow (a sympathy which is supposed to be unpleasant to undergo) could ever yield altruistic concern, rather than irritation or even hatred. Hume tries to address this worry by distinguishing between “extensive” and “limited” sympathy. According to the first description he offers of the difference between these two forms of sympathy, only extensive sympathy “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; possible, probable or certain.” Based on this description, the distinction between extensive and limited sympathy seems to be that the former enlivens our representations of more of the other’s emotions. We are given no reason to think that the expansion of our sympathy’s scope will change the basic nature of our relation to those emotions. The difference between the two forms of sympathy looks like a quantitative, rather than a
qualitative one. Perhaps the thought behind the claim that extended sympathy will cause us to love even those who suffer is that extended sympathy leads to an enlivening and conversion of presently suffering people’s future hopes and joys. With the enlivenment of these future feelings, which are much more pleasant to sympathize with, the balance of the pleasure and pain we take in sympathizing with suffering people shifts, and we come to love them.

How ought we to assess this account? First, the good news: it does not make the emergence of love and benevolence contingent upon any epistemic mistakes on the sympathizer’s part. Even if one has qualms about calling the ideas about others’ emotions that the sympathetic process furnishes knowledge, this first account of how sympathy yield motivation at least does not render the story about empathy’s epistemic “bootstrapping” more problematic. The motivation the sympathetic mechanism furnishes is clearly compatible with our maintaining a clear view both of the basic fact that the other is another and of the particular contents of the other’s inner emotional life. Indeed, it is partly because we maintain a clear view of the other qua other that we are led to love or resent them as the causes of our pleasure or pain. On the motivational front, the account also effectively avoids one obvious worry about “mere contagion” accounts of empathy’s motivational force. It has been repeatedly pointed out that the claim that we are motivated to eliminate the source of contagiously contracted empathic pain (along lines like: the causal source of my pain is his pain, so I should alleviate his pain) is suspect, because our pain really motivates us just to stop the pain, and we can stop pain without eliminating its original cause. An analgesic, the thought goes, could end our discomfort just as effectively as the relief of the other’s pain. Hume’s “joy of sympathizing” account isn’t affected by this objection, since the act of taking a pain-relief pill won’t necessarily relieve one’s resentment that another has (inadvertently) caused one to feel uncomfortable.

13 Nichols (2004): 36 is one of many to make this point.
Resentment and love are necessarily other-directed. They motivate us not (or not just) to change our own inner states, but also to do something to or for the other.

That said, the joy of sympathizing account is quite plainly both inadequate and implausible as an explanation of how we come by our altruistic sentiments. It implies that the most miserable of creatures, one with a fully hopeless future, will naturally be the object of our resentment, and cannot be someone whose good we could come to desire via sympathy. Certainly it is less fun to vividly picture or feel agony and desperation than joy and contentment, and it is true that we sometimes shy away from emotionally engaging with others’ struggles out of a desire to spare ourselves the discomfort. But the claim that we love and care for others insofar as they are a source of our own pleasant sensations simply does not seem to conform to our experience. Nor is that orientation toward others the one that we might have hoped sympathy could explain. The account explains why we are motivated to alleviate others’ pain, but a motivation to alleviate others’ pain is not necessarily benevolent. Wanting someone’s happiness because and insofar as their own experiences have furnished the occasion for our own joy does not amount to real benevolence, because loving someone is not the same thing as loving what someone makes possible for you. What we really care about, on this picture of Hume’s, is feeling good. We care for others just because they furnish the means to that end. As such, our orientation toward others is still hopelessly self-centered.

One notable feature of the “joy of sympathizing” account is that the sympathizer is supposed to be moved to love not directly by the enlivened passion they experience, but rather by the pleasantness or unpleasantness of the experience of the passion. That distinction can be obscured by a focusing on sympathy as it applies to others’ pain or pleasure, simpliciter. To bring out the point, take the case of someone who sympathizes with an experience of grief. Perhaps the sympathizer is moved in a
way that is characteristic of grief, but according to the joy of sympathizing account, that motivation (the desire to weep, to rend one's clothes, etc.) is not directly responsible for one's love or resentment. It is, rather, one's discomfort at feeling grief that inspires the relevant other-oriented concern. So, we love someone because we are undergoing a pleasant experience, that of feeling joy, thanks to them. Or we hate them because sympathizing with their anger is painful. One might wonder whether this is where the “joy of sympathizing” account goes wrong. Perhaps the source of benevolence should be located more directly in the sympathetic passion itself. One of the remarks embedded in the above passage suggests that Hume was at least somewhat attracted to this alternative possibility, which forms the core of what I will call the “identical motivation” account.

5. The “identical motivation” account

Hume notes that “[t]he 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.” This talk of “depth,” and of “entering into” others’ interests, seems to suggest a different view of the division between extensive and limited sympathy. Perhaps the proposal should be understood as something like this: when the sympathizer’s engagement with the other’s passions is shallow, she is, as it were, alienated from the sympathetic passion she experiences. She feels it in some sense, enough so that she experiences those passions as pleasant or unpleasant, but not in a way that maintains the passions’ usual connections to motivation. Conversely, when the sympathizer’s engagement with the other’s passions is deep, she fully takes on board the other’s passion, such that it moves her in just the same way that the original version moved the other.

The idea that Hume at least partially or intermittently embraces this latter possibility receives further support from his claim that when others’ sentiments affect us, “they operate upon us, by opposing and
increasing our passions, in the very same manner, as if they had been originally derived from our own temper and disposition.” Recall, while Hume doesn’t strictly attribute intentionality to passions, he is willing to allow that they direct our attention in particular ways. So, if the sympathetic passion directs attention in just the same way as the original passion, then if my fear is directed toward the barking dog, the one who sympathizes with me will end up also fearing the dog, and being motivated in just the same way that I am.

On this second account, how is being motivated in “the very same manner” supposed to relate to benevolence, exactly? Hume doesn’t give us much to go on, here, but fortunately for us this latter approach to accounting for the empathy/benevolence link has been developed in the recent literature. Let us consider Michael Slote’s proposal along these lines. Slote explains the working of empathy through the example of a father whose daughter is interested in stamp collecting. The father catches the daughter’s enthusiasm, “takes it in by a kind of empathic osmosis.” He consequently becomes enthusiastic, not in sense that he develops a generally enthusiastic mood, but in the sense that the “enthusiasm has the same intentional object as his daughter’s, namely, stamp collecting (there is some ambiguity or leeway here as to whether he starts wanting to help her collect stamps or starts wanting to collect for himself or both).”¹⁴ This enthusiasm is, like the original attitude it is copied from, the genuine article. The father now actually cares about the hobby. The same pattern of transmission will go for people’s other pains and pleasures, loves and hatreds. And the fact that the mechanism has this general application means that it can serve as the bedrock of all benevolence:

If someone feels pain and is distressed about it, then they automatically, ex vi termini, count as motivated. That’s just what distress means. But then consider someone who empathizes with, who automatically takes in, the other person’s distress at their pain. This means feeling

distressed oneself about their pain, and again *ex vi termini*, this constitutes motivation to alleviate that person’s pain. Which is what we mean by altruistic or benevolent motivation.\(^\text{15}\)

This “identical motivation” account obviously has one major advantage over the “joy of sympathizing” account. It does not make our care for others’ suffering and joy secondary to and dependent on our interest in feeling pleasant emotions or sensations and avoiding painful ones. Other people don’t (typically) care about their own emotional experiences in that way, after all. When they are afraid of some danger, for instance, people are not principally and primarily concerned with how unpleasant their experience of fear is. Rather, they are concerned with and motivated to avoid the menacing stranger or rabid raccoon. And since the empathizer is presumed to be motivated in just the same way as the other, she will not come to care about the other’s passions just as means to the end of avoiding painful emotions or sensations. If the other’s passion is principally concerned with some circumstance in the world (such as the presence of the rabid raccoon) then the sympathizer’s passion will be as well. However, the identical motivation account runs into other difficulties. Specifically, makes it look like the care generated through empathy doesn’t actually end up being properly concerned with the other person and/or hinges on a distortion of the others’ attitudes.

Many commentators have insisted that Hume’s account of how we end up being motivated to aid others is not in the least bit egoistic, since the emergence of our desire to aid does not involve us caring about others’ joys or sorrows in virtue of their “belonging” to us.\(^\text{16}\) I grant that according to the way Hume’s (and Slote’s, and other similar) accounts are supposed to work, the sympathizer’s concern is not dependent upon an explicit characterization of the other’s pains and sorrows and joys.


\(^{16}\) Although a number of commentators have maintained that the *Treatise’s* take on sympathy has a whiff of egoism about it, it is more common to find that interpretation dismissed as an obvious mistake. Postema says that interpretation is “misplaced,” because “Although pity depends on sympathy, and so on an implicit resemblance between the pitier and the pitied, it is impersonal. The self is not in focus” (2005): 273. See Ardal (1966): 69, and Raynor (2006): 241 for similar claims.
as belonging to her, the sympathizer. But that does not prevent the “identical motivation” account from looking like it doesn’t adequately get us outside of ourselves. To show why not, let me focus on the case that has most interested proponents of accounts like “identical motivation,” namely our sympathetic engagement with others’ pain-oriented distress.

Slote says that we are when we feel distress at pain, we are motivated to alleviate it. And, as Thomas Nagel has pointed out, an aversion to pain need not involve the idea that the pain is mine. Nagel describes the relation to suffering this way:

The sufferer’s reaction is very clear. Of course he wants to be rid of this pain unreflectively—not because he thinks it would be good to reduce the amount of pain in the world. But at the same time his awareness of how bad it is doesn’t essentially involve the thought of it as his. The desire to be rid of pain has only the pain as its object...If I lacked or lost the conception of myself as distinct from other possible or actual persons, I could still appreciate the badness of the pain, immediately.17

We ought to grant to Nagel, and to all the commentators who insist that Hume’s account is not covertly egoistic, that my immediate aversion to pain does not depend on my representation of a pain as my own, and that when you experience a copy of my aversion thanks to sympathy, you need not think of the pain toward which you now feel averse as your own. Nagel describes the typical object of the sufferer’s aversion as “the pain,” in contrast to “my pain,” and that seems plausible enough. But it is important that we understand “being averse to ‘the pain’” in the right way. The fact that I need not think of the pain I experience as my own does not entail that what I am averse to is some thing, pain, whose aversiveness for me is independent of whether or not I am undergoing it. Fundamentally, what I as a suffering person normally hate is not this or that pain, but rather being in pain.

That contrast may not seem intuitive. Let me try to bring it out with an analogy: it is like the difference between loving and hoping for some particular act of fishing, on the one hand, and loving and hoping to fish, on the other. Neither of these two kinds of attitude requires us to invoke a concept of self in order to characterize it: in that sense, at least, the latter is no more “selfish” than the former. Still, there is an important difference between the two. In the case of my hoping for some act of fishing, it doesn’t strictly matter to the satisfaction of my hope that I be the one to do the fishing. Perhaps I am, as a matter of fact, the only person who can perform this particular act of fishing, but if it could somehow be accomplished in another way, I would be satisfied. The same is not true for the hope to fish. This hope is not identical to the hope that I will fish, but what it has in common with that still other possible desire is that only my doing the fishing will realize it. Similarly, I want to suggest, my immediate, primitive aversion to pain is really an aversion to being in pain, and only my ceasing to be in pain will be appropriately responsive to the desire at the heart of that aversion.

I do think that a characterization of our aversion to our own pain that does not allow that we hate being in pain, not just “the pain,” will distort the nature of our relation to our own experience. But I am not going to offer proof for that claim here (I invite you to check whether it lines up with your own experience). Rather, I will simply show that whether one understands our primary aversion as (1) hating being in pain, or rather as (2) hating the pain, where this amounts to something aside from hating being in pain, the “identical motivation” account falls apart.

First, consider what follows if I am right about the nature of our primitive relation to our own pain. If the two-stage sympathetic process is to furnish me with an accurate conception of the other’s emotional state, then the idea of the other’s mental state which is to be enlivened in me must exactly reproduce the original state. And if the other’s original state is distress at being in pain, then the idea to
be enlivened in me is an idea of distress at being in pain. Enlivening, as we have seen, changes only the
vivacity, and not the content, of this idea. So, after the second stage of the sympathetic process, I end
up feeling…distress at being in pain. And distress at being in pain will motivate me only to relieve the
pain that I am in.

Can a motivation to relieve the pain that one is in be a form of altruistic motivation? A defender of a
Humean bridging account might, at this point, be inclined to respond that depending upon how we
identify the pain the empathizer hates being in, the “mere” motivation to relieve the pain that I am in
could get us all the other-oriented concern we could possibly want. This is the strategy adopted by
Schopenhauer, who is if anything more deeply committed than his hero Hume to the thesis that
empathy is the sole, “miraculous” means by which we are able to overcome a primitive egocentrism so
intense that we typically see ourselves as the unique center of the world, the “owner of all reality.”18

Schopenhauer follows the “identical motivation” thread in Hume’s account, helping himself to the
thought that I can only come to be moved immediately by another’s well-being if “directly suffer along
with him, feel his woe as otherwise I feel only mine, and so will his well-being immediately as I
otherwise will only mine.”19 As for what sympathy involves, he has this to say: “I must rebuke the error
of Cassina, who holds that compassion comes through a momentary deception, of fantasy, as we
ourselves substitute ourselves in the place of the sufferer and then, in our imagination, take ourselves to
be suffering his pains in our person. It is not like that at all. It is in his person that we feel the
distress.”20

Schopenhauer wants to grant that if sympathy worked in such a way that my sympathy with my neighbor’s pain in her arm would end in *my* hating the pain in *my* arm, then sympathy could not be a means of overcoming primitive egocentrism. But I take it that his key thought is this: to characterize the work of sympathy properly, we need to allow that the pain the original sufferer hates being in is *this pain, here*, and that the sympathizer ends up hating being in *that very same pain*. So, for example, if you hate being in pain, and the relevant pain is the pain in your arm, then if I sympathize with you I will end up hating being in the pain that is in your arm. My sympathetic response will accordingly be to seek the relief of the pain that is in your arm, and this, Schopenhauer thinks, is enough for compassion or benevolence. In fact, one might further insist, it gets us everything we could possibly want: I as a sympathizer end up caring about the pain in your arm just as fiercely and intimately as you do.

If sympathy works as Schopenhauer thinks it does, then it is true that I, the sympathizer, will be motivated by the care it generates to do things that *look* compassionate, such as tending to the wound on your arm. But can the upshot of Schopenhauerian sympathizing be actual compassion, or only a superficially similar imitation of it? The locus of the empathizer’s concern, on Schopenhauer’s picture, is still an experience of pain he is undergoing, even if the pain is “in” the body of someone who is (empirically) other. Indeed, it has to be, or else the concern cannot retain the ferocity and intimacy Schopenhauer is after. However, the kind of attitude that we tend to refer to as compassion, benevolence, or other-directed care requires a robust distinction between self and other, and between the other’s pain and my own. The concern that Schopenhauer’s sympathy can produce, conversely, relies upon and emerges from a kind of colonization of the other’s pain. Max Scheler memorably describes what he sees as the flaw at the heart of proposals like Schopenhauer’s: the idea that empathy rightly causes us to care about others’ woes as our own makes a virtue of mental parasitism, he writes,
and would have us be “spiritual vampire[s] who “fill [our] empty bell[ies] with the experience of others.””

It is true that depending upon whether we think of the ultimate object of the sympathizer’s concern and aversion as “being in this pain, in my arm here” or as “being in this pain, in her arm there,” the facts about what has gone wrong on the epistemic front will be different. In the former case, one’s sympathetic concern seems to be directed at a delusive pain, one that is either not real or that at the very least has a strange sort of existence, having been conjured up by the sympathetic process alone. Its maintenance would seem to require a bizarre neglect or forgetting of the fact that one is not really hurt. In the latter case, the pain the sympathizer is concerned with is real in a perfectly normal sense; it exists, and would exist even in the absence of the sympathetic effort. The target of the sympathy is really hurting. But the sympathetic response does seem to involve a different epistemic mistake.

To bring it out, consider the following case. We are acquaintances, and we meet to chat about our lives over a cup of coffee. You are recovering from a painful surgery, and describe to me how awful the experience has been. I frown, and say, “That’s awful.” So far, so good. That response seems appropriate. Then I say, “You are right to be frustrated.” That’s also fine. But then, my face crumpling, I exclaim: “Ah! Ah! It’s so hard to cope with this pain in your hip!” Here, we want to say that something has gone very wrong. The pain in your hip just isn’t mine to hate being in. I’m not wrong about the reality of the pain, or where the pain is, but I am wrong about how I could possibly be entitled to relate to it.

---

Because the errors on the epistemic front are different, the respects in which these two possible outcomes fail to actually amount to other-directed care are different. In the former case, the case in which the sympathizer ultimately ends up hating “being in this pain, in my arm here,” the sympathizer’s range of concern is not expanded to harms outside of her (physical) self at all. In the latter case, the range of concern is dramatically extended in one sense; pains in far-off places are experienced as awful and to be relieved. However, even if hating being in some far-off pain does not require the sympathizer to conceive of the pain she is in as hers, it still seems to involve a usurpation of the other’s experience, rather than a concern for the other’s experience as such.\textsuperscript{22}

We have been considering what the outcome of sympathizing would be, according to the “identical motivation” interpretation of Hume’s account, if our most basic attitude toward pain were to involve hating being in it. It turns out that in that case, it does not matter whether the feeling generated in the sympathizer is simply a hatred of being in pain, or more particularly a hatred of being in that pain over there. Either way, the attitude in question will not really be concerned with or directed at the other, as an other. As we have seen, the operation of sympathy on this interpretation would necessarily entail some confusion on the sympathizer’s part. So, what we have is a double fault: sympathy would seem to get things wrong on both the epistemic and the motivational front.

Now, let us suppose that I am wrong. In reality, let us say, it will typically be correct to characterize the other’s basic suffering attitude as involving a hatred of “the pain,” where this is not a matter of hating being in pain but rather, somehow, just a hatred of the pain itself. If that were true, then my accurately

\textsuperscript{22} Schopenhauer himself would reject the notion that any “usurpation” is afoot here, since he ultimately denies that there is any real difference between self and other. For Schopenhauer, the idea of a genuine plurality of persons is an “illusion” (2010): 249ff. Schopenhauer’s argument for this “no difference” view is a weak one: he reasons from the Kantian thought that our individuation of beings relies upon concepts of space and time that are ideal, to the conclusion that there must be no individuation in the realm of things in themselves. As Paul Guyer bluntly but not unfairly points out, this is an “enormous fallacy” (2012): 413. If the wholesale dismissal of the highly intuitive notion that people are genuinely separate is to have be at all tempting, a much stronger argument than Schopenhauer’s will be required.
copying and enlivening the idea of her passion will not result in my curiously folding in on myself, and becoming motivated just to relieve my own suffering, or upon a colonization of another’s suffering. Suppose the target of my sympathy hates some particular pain \( Q \) (not having or experiencing \( Q \), just \( Q \) itself). Then, if I properly sympathize with her (we might think) I will also come to hate a particular pain \( Q \), and be motivated to eliminate it.

The possibility of getting to non-instrumental concern for a somebody else’s pain (\( Q \)) in the way just suggested can only constitute a hollow triumph for the “identical motivation” account, though. Why? Well, this last story about how we get to the requisite motivation makes it possible for sympathy to resolve the problem of motivational egocentric primacy only by effectively denying the picture of primitive self-concern that got the problem up and running in the first place. If it were the case that our primitive concern for the weal and woe that happened to be our own were impersonal, which is to say if the concern were not about “our pain” or “being in pain,” but were merely about pains whose badness could be characterized independently of their relationship to us, then we would be left with the question: why should we think that empathy would have anything to offer here in terms of getting us to care more about other people? We will have effectively solved the problem, but left it unclear what work, if any, there remains for empathy to do.\(^{23}\) Empathy’s promise seemed to be that it would allow us to bootstrap up from the special relationship we have to our own inner life, our own pains and pleasures, in order to arrive at knowledge of and care for others. But the “solution” we are considering here doesn’t look like a way of vindicating that promise.

\(^{23}\) Interestingly, Adam Smith obliquely points to this consideration in remarking that when it comes to our attitudes about things whose goodness we experience primarily as being objective or impersonal, such as art and literature, there is no special role for empathy to play in fostering the relevant care or appreciation: when it comes to “the beauty of a plain, the greatness of a mountain, the ornaments of a building, the expression of a picture, the composition of a discourse, the conduct of a third person, the proportions of different quantities and numbers, the various appearances which the great machine of the universe is perpetually exhibiting,” Smith contends, “We both look at them from the same point of view, and we have no occasion for sympathy, or for that imaginary change of situations from which it arises, in order to produce, with regard to these, the most perfect harmony of sentiments and affections” (TMS 119).
It is worth pointing out, too, that while the appropriateness of describing our basic motivation relative to our own pain as impersonal might be something on which respectable moral psychologists could disagree, the same will not go for many other attitudes. If we as sympathizers correctly represent others’ humiliation or pride, for instance, our enlivening of those representations will plainly turn us back towards ourselves. We will feel humiliated, and be motivated to hide ourselves away. Or proud, and motivated to strut. Those motivations will be far off indeed from other-directed care, because they seem to be all about me. This in itself is a disappointing result, and one that I doubt Hume would have been willing to embrace.

6. A “transformative motivation” account?

As we have seen, attempts to flesh out Hume’s treatment of empathy quickly end up tying us in knots. The enlivening of an accurate idea of the another’s passion simply does not seem to yield an emotion that has the right kind of other-orientation, nor does the sympathizer’s interest in having or avoiding experiencing a copy of another’s emotion seem to be a viable source of the kind of care for others we were looking for. What is more, with the exception of the “joy of sympathizing” account, the routes out of primitive egocentrism we have pursued have all involved epistemic error or distortion on the part of the sympathizer: the sympathizer is required to believe she is experiencing pains in other bodies and/or to lose sight of the distinction between her own and others’ situation and experiences. These errors are, it is worth empathizing, strange ones. It is of course true that they are ones we sometimes make. The kind of parasitism Scheler decries, in particular, is one way of doing what we call “living vicariously.”

Interestingly, on Nel Noddings’ conception of mature empathy as a “sharing of motive energy,” empathy involves a “displacement,” and she notes that “when this displacement occurs in an extreme form, we sometimes hear parents
corrected and avoided, rather than an activity that could plausibly generate and sustain our basic interest in and knowledge of others’ inner lives. When we find ourselves attributing confusion or pathology to ordinary people engaging in an ordinary way of relating to other people, we ought to reconsider, and try a different tack.

I now want to turn to one newer strategy for describing empathy’s bridging capacity that retains some elements of Hume’s approach, and that crops up repeatedly in the literature. If the feeling at which one arrives via the sympathetic process is just a copy of the other’s original feeling, the feeling of hating being in pain, then it can obviously not itself constitute an altruistic attitude. But it seems we can’t resolve the difficulty by positing that the primitive concern for pains and pleasures with which we begin is essentially impersonal either, on pain of making the work of empathy look merely redundant. What we need to do, then, one might think, is to posit a kind of transformation of the attitude empathy originally generates in us, such that it becomes altruistic.

Hume himself was adamant that the feeling we end up with at the end of the whole sympathetic process is an exact copy of the original passion. He claimed that the mind’s associative mechanisms “entirely convert an idea into an impression, and convey the vivacity of the latter into the former, so perfectly as to lose nothing of it in the transition” (SBN 320). However, some philosophers have suggested that we can get to genuine concern for the other without distorting the original, self-centered nature of the other’s attitudes so long as we allow that one’s sympathetic emotion undergoes a shift in orientation. Heidi Maibom articulates this suggestion as follows: “By replaying emotions in ourselves, our attitudes and motivations towards others change. They [the replayed emotions] move us similar to the way our own desires and emotions do…Emotions have a motivational component. This

__________________________

speak of ‘living for’ their children” (1984): 33. So, for Noddings, empathy is, if not actually vicarious living, at least on the same spectrum as it is.
component is transformed at the end of a simulation. Our motivation to act becomes altruistic because it has as its object someone else’s welfare, concerns, and projects.”

Because “empathic upset is still upset,” she concludes, “Whatever motivations are connected with being upset will be connected with being empathically upset.”

Based upon this rather minimal description of the empathic transformation of “upset,” we can get some grip on what the critical shift is supposed to be. The thought here seems to be that while I empathize, I initially feel distressed at being in pain, but then I remember that my distress has an atypical object: I am not myself actually hurting. After recalling this fact to mind, I still have all the motivational “oomph” supplied by my feeling distress (what Hume would call the “vivacity and forcefulness”), but I change the object of my distress so as to better conform with the actual facts of the situation; to wit, that there is only one person who is genuinely experiencing hurt, and it is not me. The distress at being in pain I experienced through empathy thus gets converted into distress at another’s pain. Maibom doesn’t say it, but presumably she would want to add that this latter distress is genuinely about and for the one experiencing the pain, whom I recognize to be an independent other (otherwise, as I’ve tried to bring out earlier, the end product won’t really be compassion, but rather a kind of “colonizing” personal distress).

Developmental psychologist Martin Hoffman offers a strikingly similar picture of conversion, albeit one that unfolds across a critical span of psychological development, rather than across the duration of an individual empathetic experience. Hoffman claims that one empathizes insofar as “one feels what the other feels or may reasonably be expected to feel in the other’s situation,” but that

primitive empathy, the empathy that we exhibit in infancy, does not involve or entail any other-directed motivation. As young infants, we cry when others cry, but we do not even have a distinct sense of self or other-hood. This primitive empathic response, Hoffman suggests, can appropriately be called “egocentric’ empathic distress,” because it elicits in the distressed individual a self-caring response. But around the second year of life, he claims, individuals gradually become aware that others have distinct bodies and also distinct inner lives, and so develop the metacognitive awareness that is necessary for the emergence of altruism. Hoffman describes the shift: “As part of the child’s growing sense of self and others as separate beings, empathic distress is transformed partly into sympathetic distress: the child continues to feel empathic distress, more or less matching the other’s feeling, but now adds a reactive feeling of sympathetic distress or compassion for the other.”

Hoffman’s basic thought seems to be that the concern we feel originally, as infants, is not for anyone in particular. It is just an aversive experience of pain, a hatred of being in it. Once I learn that there are others in the world, though, my initial distress is supposed to undergo a transformative reorientation.

Versions of this transformative picture feature in some of the most influential psychological work on the empathy/pro-social behavior connection. Nancy Eisenberg, another leading figure in the field, also posits that “empathy, if above some minimal threshold, is likely to evolve into sympathy,”

---

27 Hoffman (2014): 80. In his 2001 book, Hoffman uses very similar language: “Once children have separate images of themselves and others, their own empathetic distress, which is a parallel response- that is, a more or less exact replication of the victim’s actual or presumed feeling of distress- may be transformed at least in part into a more reciprocal feeling of concern for the victim” (87).

28 Daniel Batson, whose name is perhaps the one most strongly associated with the hypothesis that there is a tight connection between empathy and altruism, used to hold that empathetic perspective taking was the most “basic construct” that enables altruism. He has since revised that position, and now holds that in “everyday life, perspective-taking lies a little downstream from valuing” other people’s welfare (2011): 44. His reasons for this revision are separate from the problems I have been discussing; he cites empirical evidence that we can adopt perspectives without caring, and care without adopting perspectives, as the basis for his change of heart.
where sympathy is understand as altruistic concern for others.\textsuperscript{29} Eisenberg and her co-author Natalie Eggum endorse Decety and Jackson’s (2004) claim that empathy has three functional elements: affective sharing, awareness of the self-other distinction, and “mental flexibility and self-regulation.”\textsuperscript{30} On Eisenberg and Eggum’s view, only the appropriate combination of these three elements yields sympathy. Affective sharing is necessary for vicarious feeling (feeling the same thing as the target individual): “individuals most likely to experience empathy may be those prone to at least moderate levels of vicarious emotion,” but affective sharing is not yet altruistic.

Among those who are prone to feeling vicarious emotion, Eisenberg and Eggum distinguish between two kinds of person, those who are good at regulating their emotions, who can “modulate their negative vicarious emotion to maintain an optimal level of emotional arousal,” and those who lack this regulatory ability. The latter group’s empathy will tend not to be converted to altruistic care, they think, because their emotional over-arousal will leave them stuck at personal distress. Empathy will only lead them to care about their own feelings. Conversely, those who can regulate their emotions in light of their awareness of the self-other distinction are apt to end up with an “optimal level of emotional arousal.” This level is turn hypothesized to “promote an other-focus and therefore [is] expected to be associated with sympathy.”\textsuperscript{31} The picture seems to be that carefully bearing in mind the fact that one’s vicarious emotion is just that will both lessen the intensity of one’s emotion and redirect the emotion away from oneself, toward the experiences of the other. And with this redirection, the emotion is supposed to become altruistic in character.

\textsuperscript{29} Eisenberg and Eggum (2009): 71. Note the remarkably clear echo of Hume in this formulation: the idea of a “threshold” of empathetic involvement recalls Hume’s distinction between vivacity of conception “confin’d merely to its immediate object” and vivacity not so confined (SBN 386).

\textsuperscript{30} Eisenberg and Eggum (2009): 72.

\textsuperscript{31} Eisenberg and Eggum (2009): 72.
All of these accounts—Maibom’s, Hoffman’s, and Eisenberg and Eggum’s—invite the question: why exactly should we think that an awareness of the distinction between self and other, and of the empathetic nature of one’s emotional experience, will transform the empathizer’s emotion into an altruistic one? It makes sense to grant that such an awareness will tend to focus our attention on the other, but it is not clear why this shift in attention would trigger the transformation of the empathizer’s emotion such that it becomes other-directed. This is not a minor oversight that can easily be corrected. On the contrary, there is good reason to think that proponents of this kind of account simply will not be able to satisfactorily explain the purported motivational shift.

To see why not, consider first what kind of response to realizing she is not herself really in pain it would make sense for the empathizer to have. If all we primitively care about, when it comes to pain, is not being in it, then it is totally unclear why we shouldn’t, upon remembering our actual, unpained situation, simply laugh off our “upset.” Far from giving us a reason to transform our pain into altruistic concern, the discovery or recollection that we (the sympathizers) are not actually in pain, that we are in fact “just” empathizing, seems to give us reason to simply disregard our aversive empathetic feelings. The posited reorientation of our aversion seems utterly unmotivated, unless one is willing to allow that it motivational orientation the conversion in question is supposed to generate.

Defenders of transformative accounts may not be very impressed with this point. The relevant task, they might respond, is not to explain how or why a reorientation of the empathizer’s pain could possibly look reasonable to the empathizer. All we need to do is elucidate the causal mechanisms by which the awareness of the self-other distinction triggers a shift in the orientation of upset, and that should be no harder to explain than any other shift in emotional orientation. Emotional upset does not disappear, but must instead be channeled along some line of association or other. This kind of thing happens all the time. The orientation of our rages, hopes, and affection alters according to
shifts in our attention, and these shifts are not always “reasonable” or otherwise aligned with our pre-existing motivations. The emergence of altruistic attitudes will be just one example among others of this phenomenon, one might think, and its full explanation is the kind of challenge to which we might reasonably hope brain-imaging and animal experimentation will eventually satisfactorily respond.\footnote{For one such attempt, see Tucker, Luu, and Derryberry (2005).}

This is an explanatory strategy we should be reluctant to accept. It casts the empathetic generation of altruism as roughly on a par with letting one’s anger at one person bleed into one’s interactions with another, or having one’s delight with a good meal transform into optimism about the political situation. That is to say, it threatens to make the care arrived at via empathy look like a \textit{mistake.}

Suppose I feel upset at my sister and I yell at my blameless brother. There doesn’t seem to be a relevantly significant difference between that outcome and feeling distressed at my own pain and (somehow) letting it bleed into concern for another’s, despite the fact that I have no antecedent motivation to care for the other.

That observation may not be universally troubling. Perhaps some philosophers will be quite happy to accept that our basic concern for others is a happy accident, one that emerges as a result of the “leakiness” of our emotional attitudes, although the proposal certainly sits uneasily with Eisenberg and Eggum’s insistence that the key to transforming empathy into altruistic care, rather than personal distress, is being sufficiently skilled at “regulating” one’s emotions. Why should we not to go along with that cheerful acceptance?

One of our strong initial intuitions about empathy, as discussed in the first chapter, is that empathy (or, at least, virtuous empathy) helps to put us in a better epistemic position, making us wiser or
more understanding. On the transformative view we are contemplating here, though, our empathetically generated care is supposed to arise through a mechanism that exploits an irrational mental tendency, namely the tendency for our emotions to take on new objects through associative “leakage.” There is a sort of tension, here. It is not a totally straightforward one; after all, the transformative picture does allow that through empathy we come to know new things, namely the nature and content of other’s attitudes. But still, the picture does pit the end of thinking and responding in a reasonable, sensitive, and clear-headed way, one that we might have thought empathy would contribute to, against the end of altruistically caring for others. According to this sort of proposal, it seems, someone who is truly, maximally mentally disciplined would not care for others as a result of empathically engaging with them. Rather, they would keep their empathetic emotional upset confined, or (better still) repress it entirely. And this result, in addition to being distasteful, sits poorly with our intuition that empathy enhances both our epistemic and our moral virtue, rather than pitting them against each other. Embracing the transformative account, then, would mean not only accepting a significant explanatory lacuna, but also surrendering our grip on one of our project’s foundational intuitions.

6. Conclusion

New Englanders are traditionally credited with the singularly unhelpful travelling advice “You can’t get there from here.” This chapter has been devoted to showing that Hume’s account(s) of how sympathy works, influential as they have been, don’t really explain how empathy can get us from a position of primitive motivational and egocentric egocentrism to a place of both knowledge of and benevolent care for other people and their inner lives. I’ve also argued that we should be dissatisfied with the transformational accounts favored by some contemporary philosophers and psychologists. Those depart from Hume in allowing for the transmogrification of empathic feelings’ orientation, but retain
other key elements of the Humean picture, including the idea that the “oomph” supplied by empathetic emotion is at the core of our altruistic motivation.

In philosophy, the recent revival of interest in empathy’s moral significance has focused largely on the hotly-debated question of whether empathy is necessary and/or sufficient for altruism. There is much discussion about the possibility (or reality) of empathetic but cruel psychopaths, or of moral but utterly unempathetic autistic individuals, since these fringe cases have seemed to many people critically important for determining whether altruism can possibly exist without empathy, or vice versa.33 Conversely, I have been trying to draw out the more basic puzzle of how we could possibly account for empathy’s making any difference to our care for others, particularly in a way that does not end up making our care look irrational. If this more basic puzzle has somewhat fallen by the wayside, I suspect this is in part because the difficulty is obscured by imprecision about the exact character of the empathetic feeling.

We have already seen that Slote does not consider it very important whether the stamp-collecting father’s empathetic enthusiasm is for helping his daughter or just for the stamp collection (Slote allowed that “there is some ambiguity or leeway here”). Other philosophers similarly describe empathetic feeling as “an affective state that is ‘in the same neighborhood’ as the target’s” or “similar to what the other being is experiencing.”34 This looseness is entirely understandable. After all, it would be silly to seek out some precise threshold of similarity to the other’s original attitude that an attitude must meet to qualify as empathetic. At the same time, though, it does tend to cover

33 To cite just a few important contributions to this burgeoning literature: Kennett (2002), Maibom (2008), Aaltola (2014), Denham (2011), Deigh (1995). See also Chapter Four, section 2 (171ff).

up the problems we’ve discussed. If we say simply that the two attitudes in question are “in the same neighborhood,” for instance, then it will be relatively harder to make out the worry that simply feeling a copy of another person’s hatred of being in pain is not equivalent to hating the other’s pain, in an altruistic fashion. But vagueness about the nature of the empathizer’s experience is not, of course, a solution to the difficulties discussed in this chapter. To briefly review why not: if the empathetic feeling is importantly different in orientation from the original, then that will generate problems for the Humean thesis that our empathetic recreation of the other’s experience allows us to know the nature and content of that experience, and if the empathetic feeling is not importantly different, then that will generate problems for the thesis that empathy does not just reinforce egocentrism, but rather supports altruism. No matter how loosely we characterize the resemblance between the original and the empathetic feeling, these difficulties will continue to be just as real.

We have traced Hume’s ideas about empathy through to some of their present-day incarnations, and we have seen the ways in which those ideas go wrong. In the next chapter, we will return to 18th century Scotland, and consider Adam Smith’s analysis of empathy. Smith’s account is not without its flaws. Still, it offers us the beginnings of a very different picture of empathy’s moral and epistemic significance, one that not only avoids the problems that we have already seen, but also reveals itself to be beautifully sensitive to the actual complexity of our attitudes about empathy.
1. Introduction

In the previous chapter, I considered ways of thinking about empathy’s epistemic and moral significance that emerge from Hume’s *Treatise*. Some of these continue to figure prominently in philosophical and psychological analyses of empathy. However, as I’ve shown, each of them is deeply problematic. Some of the accounts we’ve considered end up making it look like empathy merely expands the scope of one’s egoism, rather than generating genuinely other-directed care. Others avoid that trap, but at the expense of flattening out the very same differences between self and other that were to supposed to make empathy needful in the first place. In this and the following chapter, I will develop an account of empathy’s significance that takes inspiration not from Hume, but instead from Adam Smith. Smith’s own account of “sympathy’s” importance for human life is eclectic.\(^1\) Particularly in the first pages of his *The Theory of Moral Sentiments* (henceforth TMS), some of his boldest and clearest assertions about empathy’s work seem to echo ideas of Hume’s. Nevertheless, Smith’s most careful observations about how we experience empathy (both as empathizers and as those empathized with) contain the seeds of an importantly different vision of the work that empathy can do for us.

As we’ve seen, Hume’s account divided the empathetic process into two stages, each with their own, discrete payoff. During the inferential phase, which does not involve feeling on the part of the empathizer, but which does involve a kind of representation of the other’s emotion as from the

---

\(^1\) Except when directly quoting Smith, I will use “empathy” rather than “sympathy” to name the phenomenon that concerns both Smith and me.
inside, we secure the relevant special epistemic benefit: knowledge of the other’s inner life. And
during the second phase, which introduces the “enlivening” force of feeling, we secure empathy’s
characteristic motivating force. Hume’s take on empathy’s significance can accordingly be summed
up with two distinct necessity theses: (1) the (un-vivified) representation of others’ emotion as from
the inside is necessary for knowledge of others’ attitudes, and (2) the vivification of such a
representation is necessary for altruistic motivation. Interestingly, this picture sits uneasily with two
widespread intuitions about empathy, namely (1) that empathy’s feeling nature matters to the quality
of the knowledge it affords, and (2) that the way empathy matters to our care for others has
everything to do with the way it changes our understanding of their inner lives.²

Taking my cue from those latter two thoughts, I will propose an alternative to Hume’s picture of
what empathy can do for us, which is summed up by what I will call the humane understanding thesis:
empathy is the source of a distinct epistemic good, humane understanding, which consists in the
appreciation of the intelligibility of others’ emotional perceptions. Humane understanding is
necessary for (and, indeed, partially constitutive of) fully virtuous relations with other people.

This chapter will be devoted to an argument that empathy affords a distinct epistemic good, and the
chapter that follows it will explore the moral value of that good. To get at the nature the good, I take
up and modify a critical insight from Smith about the connection between empathy and
understanding. Smith has been accused of “develop[ing] no significant theory of mind at all,” and it

² These intuitions are sometimes reflected in the philosophical literature. Justin D’Arms approvingly mentions the idea
that way empathy shapes our care for others has everything to do with the fact that it “produces ‘emotional
understanding’ of the plight of others, rather than mere ‘intellectual understanding’” (2000): 1469, although he does not
himself defend this claim. See Barnett and Mann (2013): 230 for a very similar formulation. Heidi Maibom similarly
suggests that empathy’s moral impact is grounded in the fact that when we empathize, “understanding ceases to be
disengaged” (2007): 162. Unfortunately, Maibom’s way of spelling this out does not sufficiently distance itself from the
Humean picture (see pages 92-93 of this text).
is true that his picture of the mind is less detailed than Hume’s. But if we read between the lines, tweak some of Smith’s formulations, and fill in his account with insights from recent work in the philosophy of emotion, we will end up with an account of empathy’s special contribution to our apprehension of others’ mental states that is coherent and conforms well with our actual experiences of empathy. To fully develop that picture, I will articulate how the understanding secured by empathy differs from that supplied by two other ways of approaching others’ inner lives, one of which involves no imaginative perspective-taking, the other of which involves no emotional engagement. It has been claimed of each of these ways that they are the actual source of the epistemic goods empathy is purported to deliver. We should not be swayed by these claims, however. Humane understanding requires the special combination of emotional engagement and perspective-taking characteristic of empathy.

One clarification is needed before turning to Smith. In asking what empathy makes possible, both epistemically and morally, we might have one of two different questions primarily in view. First, there is the question of how a lifetime of empathizing and being empathized with shapes our relations with others, and whether or how a life devoid of emotionally-charged imaginative perspective taking would be constrained or distorted. Second, there is the question of what someone who failed to empathize with a fellow person would be missing out on, even if that someone had an ordinary psychological developmental history, with plenty of past experience as an empathizer and as a target of empathy. This latter question is the one I will address first in offering my own Smithian account of how empathy matters. While the question of where we would be without a history of empathy is an important and fascinating one, it is also considerably more complex than the question of what is lost in the case of an individual failure to empathize, and so it is best approached after we

---

have some grip on what we (normal people generally capable of and experienced with empathy) fail to know or understand when, for instance, we “happen to be employed about other things,” or “are in a grave humor,” or for some other reason just can’t or won’t emotionally work our way into another person’s perspective.4

2. A shift away from Hume

To get a grip on Smith’s account of empathy, it will be helpful to begin by highlighting two important respects in which Smith’s view immediately departs from Hume’s. The first is a difference in the characterization of what empathy involves. Smith introduces the notion that the effort to adopt the other’s perspective is a critical part of empathy. The second difference pertains to the characterization of what empathy does for us. Smith, unlike Hume, does not hold that empathy is the means by which we gain knowledge of others’ inner states. These two differences, which are closely tied to each other, are critically important to the alternative vision of empathy’s ultimate significance that we will extract from the TMS.

On the first page of the TMS, Smith characterizes empathy as the “conceiving [of] what we ourselves should feel in the like situation” (TMS 9). The reference to a “like situation,” here, already serves to set Smith’s view apart from Hume’s. As we’ve seen, Hume’s descriptions of empathy’s mechanics tend to focus on the associative connections between certain behaviors or expressions and particular feelings. By contrast, Smith puts our ability to imaginatively assume others’ places front and center in his characterization of empathy. For Smith, facts about how the world is for the other, not facts about how the other is actually feeling, are what principally shape our empathetic responses.

4 These examples of conditions when we might fail to empathize are from TMS 17 and 18, respectively.
To demonstrate the appeal of this shift, Smith lavishes particular attention on instances in which in which our empathetic feelings seem to be elicited by facts about the other’s circumstances, even though it is not possible that the other actually feels what we feel. We feel misery when we imaginatively take up the wretched position of an insane person who is cheerfully unaware of his own condition, and horror when entering into the perspective of the dead, who are themselves insensible to the pain of being “shut out from life and conversation” (TMS 12). These are the kinds of cases where Smith’s shift away from tying empathy to the perception of others’ emotional responses is supposed to show its advantage most clearly. The insane person laughs joyfully, the dead person exhibits no expression at all. Consequently, Hume would be constrained to say about the former case that we should expect the insane person’s cheer to elicit a joyful empathetic response. When it comes to the latter case, it is not clear that any empathetic response to the dead is possible at all, according to Hume’s view. But, Smith thinks, those conclusions plainly fail to align with our actual experience as empathizers.5

Smith also brings his claim that “sympathy…does not arise so much from the view of the passion, as from that of the situation which excites it” to bear on cases that seem to tell more in favor of the Humean view, where it might seem natural to say that one “catches” a passion, desire, or action tendency without any thought of the other’s situation. For instance, Smith points to an audience’s tendency to “writhe and twist and balance” like the dancer on a slack rope (TMS 10). Smith insists

5 At one point, Hume does mention that the sight of a typical cause of an emotion can, via an associative link, vivify a representation of the relevant emotion. He writes: “Were I present at any of the more terrible operations of surgery, it is certain, that even before it begun, the preparation of the instruments, the laying of the bandages in order, the heating of the irons, with all the signs of anxiety and concern in the patient and assistants, would have a great effect upon my mind, and excite the strongest sentiments of pity and terror”(SBN 576). However, this line of association is not the one that figures in Hume’s other descriptions of the empathetic process, which focus on the association of observed expressions of emotion with vivified representations of emotions. It is not clear that Hume would ever allow the associative link between an emotion and its typical cause to override, in the empathizer’s mind, the associative link between an emotion and an expression of it.
that attention to the other’s circumstances is essential even here. The audience writhes and twists not just because it is what they see the performer do, but because that is what “they feel they themselves must do if in his situation” (TMS 10). The same goes for expressions of suffering or happiness. The mere sight of others’ outward manifestations of pain or pleasure might seem to spark an immediate matching response in us, without any reference to their situation, but this can be explained with a creative redeployment of Hume’s own associationist psychological principles: “If the very appearances of grief and joy inspire us with some degree of the like emotions,” Smith claims, “it is because they suggest to us the general idea of some good or bad fortune that has befallen the person in whom we observe them” (TMS 11).

Smith does not announce his second departure from Hume in such clear terms. It is, rather, one that simply becomes clear in the course of his articulation of his position. Hume held that the unique means by which we come to know others’ inner lives, including their passions, is via the first phase of the empathetic process: we observe another’s emotional expression, and thanks to our mind’s associative mechanics, the imagination brings to mind a very weak copy (that is to say, an idea) of a passion associated with that expression. Smith, by contrast, must allow that we can know how another person is feeling without at all initiating the empathetic process.

Why? Well, for Smith, as we have just seen, the empathetic response is dictated by the empathizer’s understanding of the other person’s situation, rather than by their expressions of emotion. In cases where there is a mismatch between the other’s situation, as the empathizer understands it, and the other’s response to that situation (as in the case of the joyful insane person living in wretched conditions), Smith is clear that our empathetic response will not correspond to the person’s actual attitudes. Therefore, looking to our empathetic response in order to learn how the other feels will
lead us astray. That consequence means that we have to accept one of two conclusions: either we cannot learn that someone has a passion that doesn’t fit their circumstances, or we must have some other way of learning what people’s passions are.

The former conclusion is obviously untenable, so Smith must accept the latter. And, indeed, throughout the TMS there are indications that Smith is something of a pluralist when it comes to our learning about others’ feelings. He allows that we can know what others are feeling by applying general rules about what kind of situation tends to incite which passion and that we can learn about another’s passions by observing the “marks” of emotion, without sympathizing (TMS 17).

Smith’s unheralded acknowledgement that there are multiple routes to knowledge of what others are feeling also jibes with our own experience. Sometimes, other people’s emotions seem to be totally laid bare, available for us to recognize simply and directly, without any need to construct an inference based on our own experience. As Wittgenstein put it: “Look into someone else’s face, and…[y]ou see on it, in it, joy, indifference, interest, excitement, torpor and so on. Do you look into yourself to recognize the fury in his face? It is there as clearly as in your own breast.” Smith doesn’t discuss it, but it also certainly looks like we can learn about others’ emotions through their sincere testimony, without empathizing with them. Finally, its also worth noting that the most typical sort of un-empathetic person, the one who naturally comes to mind when one is asked to picture such a

---

6 Wittgenstein (2009): *Zettel* 220. As mentioned before, this phenomenological point also lies at the heart of Max Scheler’s account of intersubjectivity. Scheler writes: “[W]e certainly believe ourselves to be directly acquainted with another person’s joy in his laughter, with his sorrow and pain in his tears, with his shame in his blushing, with his entreaty in his outstretched hands, with his love in his look of affection, with his rage in the gnashing of his teeth, with his threats in the clenching of his fist, and with the tenor of his thoughts in the sound of his words. If anyone tells me that this is not ‘perception’, for it cannot be so, in view of the fact that a perception is simply a “complex of physical sensations”, and that there is certainly no sensation of another person’s mind nor any stimulus from such a source, I would beg him to turn aside from such” (2008): 6.
character, is not one who has literally no idea how others feel, but rather one who knows what his companion feels but is left totally cold by it. The fact that Scrooge knows that his clerk is miserable should not compel us to conclude that he has empathized with him.

The claim that Smith (rightly) allows for un- or pre-sympathetic ways of knowing others’ inner lives admittedly does sit uneasily with the first page of the TMS, where Smith boldly states that “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 9). So long as we accept that knowledge of “what other men feel” requires having some idea of “how they are affected,” which seems uncontroversial, this declaration of Smith’s entails that no knowledge of the others’ emotions is possible in advance of empathy. It is important to bear in mind, though, that the opening paragraphs of the TMS stand in a rather peculiar relation to the rest of the text. In one respect, they point away from Hume; they introduce the idea that “the situation” is what really matters for empathy. At the same time, though, they sound much more like Hume than all that comes after them, as far as empathy’s epistemic and moral significance is concerned. This first claim of Smith’s about empathy’s epistemic significance is never repeated, nor is the claim that opens the TMS, which identifies empathy as the source of our concern for others. And the incompatibility of the above declaration with Smith’s own examples (e.g. our awareness of the insane person’s joy), which are so important to his getting the upper hand over Hume, gives us reason to treat it as something less than a careful and accurate reflection of Smith’s considered view.

It is also worth noting that a weaker version of Smith’s first description of empathy’s epistemic significance may be compatible with the admission that there are multiple routes to knowledge of another’s emotional states, at least some of which are independent of empathy. It might be that
without empathy, there will be inevitably be *something* deficient about our grasp of the other’s emotional state. After all, it stands to reason that one can have an idea of how people are affected, such that one can accurately attribute emotional states to them, while still failing to know or understand significant features of that mental state. To take some obvious examples: I might know that another is sad in virtue of their testimony, without yet knowing what they are sad about, or how their sadness feels to them, or how that sadness colors their view of their future…and so forth. The question to be asked, then, is this: if empathy isn’t epistemically special in virtue of being the sole route to securing *any* idea of others’ emotional states, then which gap(s) in our knowledge of others’ emotions is empathy particularly suited to fill in? Are any of those gaps just mentioned especially or even uniquely remediable via empathy? Or is the gap that that empathy can overcome one we have not yet mentioned?

To begin to answer these questions, we need to see how the two differences between Hume and Smith discussed in this section shape the whole Smithian story about how empathy works. Considering the details of how a typical instance of sympathizing unfolds, on Smith’s view, will help to make this clear. It is important to note at the outset that Smith’s account of what empathy involves is vague or equivocal in several critical spots. It will therefore be necessary to supplement the account in order to secure a complete picture of empathy. In the places where we need to do this kind of work, I will propose what I take to be the most plausible of the possible ways of rendering the account more precise. Once these details are filled in, we will end up with an account of empathy’s functioning that is appealing and coherent. By and large, Smith gets empathy right. There are points, however, where Smith’s vision of empathy needs not just elaboration but also correction, and I will note these as I go.
3. An instance of Smithian empathy

Smith constructs his theory around a rich diet of cases in which people succeed in empathizing, or fail to do so. Most of his examples concern some form of suffering, so I will consider an example of this sort. Smith mentions the case of a parent whose child is recently deceased. Suppose an observer finds himself face to face with this parent (a mother, let us say), sitting in grim silence with her shoulders hunched. How will the observer’s sympathetic engagement with the mother unfold? First, the observer will immediately detect a deep sadness in the mother, embodied in her downcast eyes and defeated posture. If the mother were not in the observer’s presence, he might instead learn of the mother’s sadness through a neighbor’s report. As we’ve seen, Smith (in contrast to Hume) allows for the possibility of multiple routes to such knowledge in advance of empathy.

Smith suggests that the mere recognition of the mother’s sadness may already cause some emotional stirrings in the observer, insofar as the fact of the mother’s sadness brings to mind a vague sense that she finds herself in a bad situation. But while knowledge of the other’s emotional state prepares the way for empathy, it is not sufficient for it. As Smith explains, “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” (TMS 11). Before the observer can really empathize with the mother, that is, he must know what the mother’s sadness is about. Here we see the other important point of divergence from Hume coming to bear on the description of the sympathetic process. For Hume, recall, no investigation of either an emotion’s intentional object or the context in which it is being expressed is necessary for empathetic vivification. The only connection one is required to make in order to empathize is the associative one between the other’s emotional display and one’s own feelings. For Smith, conversely, learning something of the mother’s position will be absolutely essential to empathy. A typical means
of learning about the other’s circumstances is to ask the question “What has befallen you?” (TMS 11).

3.1. Setting the imaginative scene

With the acquisition of knowledge about the other’s situation, the observer can begin to undertake what Smith calls “enter(ing) into” another’s situation, or “bringing the case home” to himself. This is the activity of imaginatively re-centering one’s perspective, imaginatively inhabiting the situation of the other. Smith stresses that the spectator’s immersion in the other’s situation will ideally be very thorough; it will involve “bring(ing) home to himself every little circumstance of distress which can possibly occur to the sufferer…adopt(ing) the whole case of his companion” (TMS 21). But the identification of that benchmark does yet tell us quite how we ought to construe the other’s “circumstance” or “case.” Does one’s “case” include only what we might call “material facts,” e.g. that this mother is childless and without independent income? Or is it also supposed to include the other’s history and preferences? How about their traits of character, or patterns of reasoning? Smith touches on these questions at one point, when he writes that “[i]n 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 317). That goes some way toward answering our questions. For Smith, imaginative changes in character may at least sometimes be part of a thorough empathetic effort.

In thinking about what we imaginatively take on when we empathize, it is important to bear in mind that empathetic engagement ranges along a scale of depth. While Smith consistently maintains that

---

7 “Enter(ing) into” may be found at TMS 9, 10, 11, 14, 23 and dozens of other points. “Bringing the case home” and variants may be found at TMS 71, 73, 75, 76, and 109, and a half-dozen other points.
the ideal empathetic engagement is one that captures within the imaginative frame every minute detail of the others’ “case,” including an exchange of “persons” as well as “circumstances,” he also allows that even an imperfect recreation of the other’s circumstances could constitute a real and valuable instance of empathizing. It is normal for an imaginary change of situation to be fleeting, incomplete, or rough (TMS 21). I can picture your “case” in finer or coarser-grained ways. Take, for instance, this situation: your tabby cat Tibbet has run away. I might empathetically imagine that Tibbet (he of the sparkling eyes and shining fur) has run away from me, or that a cat of mine has run away, or that a pet has run away, or even just that something of value has disappeared from my life.

Smith supposes that the most perfect empathy will arise from an imaginative immersion in the other’s perspective that is maximally elaborated with fine-grained detail. In an interesting discussion of romantic love, however, Smith does point out that it is often harder to empathize with another when his situation is more precisely specified. If “our friend” is in love, Smith writes, “we never think ourselves bound to conceive a passion of the same kind, and for the same person for whom he was conceived it” (TMS 31). We do not think ourselves bound to do so in part because it is something we cannot do—our imaginations do not extend to entering into the “eagerness of emotion” that the lover feels for the particular object of his attachment. We cannot imaginatively go along with a picturing of some particular person as the object of affection, but we are nevertheless able to empathize with the general condition of being in love (TMS 32). Smith does not explicitly acknowledge it, but his understanding of the aim of empathy seems to point in two different directions when it comes to the detail and breadth of what we are to imagine. On the one hand, restricting my attempts to imagine to a coarse-grained, limited characterization of your situation limits the potential for me to arrive at an empathetic perspective that matches yours precisely, in all its richness. But on the other hand, this kind of restriction may often be the best hope for securing
some kind of perspective match, even if it is a relatively shallow one.

In the next chapter, I will provide some reason to think that a maximally rich, fine-grained perspective match is not always the end toward which our empathetic efforts should be oriented. Contra Smith, imaginative engagement with a more coarse-grained characterization of your “case” may in a given case be the ideal empathetic response, rather than a sort of second-best. But this somewhat different characterization of our ends in empathizing will be compatible with the acknowledgment both that (1) Smith is right about the general orientation of our perspective-taking efforts, that what we are after is a match with the other person’s point of view, and that (2) this can sometimes be expected to involve adjusting our imaginative efforts in line with what we know of the other’s character. When we empathize, we are looking for a grasp of how the other person experiences their situation, a grasp that is, as it were, from the inside, and that grasp just is the perspective match we are able to secure.

3.2. Empathetic emotion

Having said something about what we are to imagine when we empathize, we can now carry on with the story of the observer’s empathy for the mother of the dead son. Suppose the observer has set the scene for himself. He imagines that he has recently lost his son, and is now childless and without income. Perhaps he also attempts to accommodate facts about the mother’s personality within his imaginative effort. He might accommodate the mother’s timorousness, for instance, by putting any particularly bold thoughts to one side, and trying to find within himself a nervous streak that he can bring to bear on his view of the imagined circumstances. What then? Well, Smith explains, “when we put ourselves in his [the other’s] case, [a] passion arises in our breast from the imagination” (TMS, 12). We respond emotionally to the imaginary circumstances we have set for ourselves. So,
the observer should feel in himself some response to the unfortunate situation thus described.

What, exactly, is the nature of this emotional response? Does the empathizer just feel a regular emotion? Or is his response somehow special, distinct from his ordinary feelings of joy, fear, and so forth? The vocabulary Smith uses to describe empathetic experience certainly suggests that actual feeling is supposed to be part of it. The mother who empathizes is affected by “pangs,” the empathizer experiences “some degree of the like emotions” (TMS 11), empathizers are “not improperly said” to share the distress of their targets (TMS 15). And it is clear from the start that the emotional experiences that figure in empathy are supposed to significantly resemble those of the target. In parallel with the emotions of the target, which arise in response to the situation in which they find themselves, the feelings of the empathizer are supposed to be responsive to “the case,” the imaginatively inhabited situation. Smith stresses, though, that the empathizer’s feeling never reaches the same “pitch” as the target’s: there is a “degree of passion” that is beyond his reach, no matter how effortful and careful his attempt to take up the other’s position (TMS 21). Smith forcefully reiterates the claim that other’s empathetic emotion will inevitably fail to be as intense as the original at multiple points in _The Theory of Moral Sentiments_.

The stress Smith places upon the difference in intensity of feeling might seem to suggest that this is the _sole_ significant dimension along which empathetic and original feelings differ. One might think that the empathetic emotion is to be understood as exactly like the original (same valence, same orientation, same phenomenology) except that it is less intense. However, in one critical passage

---

8 Including TMS 9, 11, 21, 22, 27. In fact, it ends up playing a key role in his account of the importance of self-command (an ability to temper the pitch of one’s own emotions is a virtue on Smith’s view, because it allows others to empathize more easily).
Smith allows that there is another respect in which empathetic and original emotional experience differ. He writes:

What they [the empathizers] feel, will, indeed, always be, in some respects, different from what he [the target] feels, and compassion can never be exactly the same with original sorrow; because the secret consciousness that the change of situations, from which the sympathetic sentiment arises, is but imaginary, not only lowers it in degree, but, in some measure, varies it in kind, and gives it a quite different modification.” (TMS 22)

What is this “quite different modification,” this difference in “kind” rather than degree? Smith never elaborates on this suggestion. Nevertheless, it will be important for us to do so. If Smith were to characterize empathetic emotions as being entirely like the emotions we normally experience in response to our own present situations, save in their intensity, odd consequences would result. Suppose I set out to empathize with a person who has just received a prestigious award. The situation is one that invites pride on the part of the winner. What would such pride amount to? Well, the feeling of pride typically involves a sense of oneself as excellent in some respect. When proud, a person conceives of herself as someone who ought to hold her head up high. And seeing herself in this light will be intimately tied to certain further tendencies of thought, behavior and feeling. The proud person might be moved to reflect at length on her accomplishments or to silently congratulate herself. She might be inclined to be a bit more self-forgiving than she has previously been, in light of what she has been able to accomplish. She might excitedly tell her friends and family about her award. Pride typically has a somatic dimension as well, though not one that is as obvious and attention-commanding as that of emotions like fear and anger: it might be characterized as a feeling of lightness.

Now, if all of these tendencies and associations are part of what it means to feel pride in the normal way, and if feeling pride in the normal way is just what the one empathizing with the winner does, then the would-be empathizer will end up in a strange position indeed. If she ends up silently
congratulating herself, reflecting on her accomplishments, or even telling others about her award, then we will be inclined to say not that she has empathized very effectively, but that she has made a peculiar mistake. We will regard her as having forgotten who won the award, or as having strangely collapsed her identity and the winner’s. It won’t help matters to say that, as the empathizer’s pride is less intense than the original, she will be only weakly inclined to do these things. A more discreet confusion is still a confusion, and we do not want to cast the empathizer as confused. Empathy is not the same thing as delusion, and we must avoid assimilating those two states.

One possible way around this difficulty is to simply conclude that empathy itself does not actually involve any real emotion on the part of the empathizer. There are two principal available versions of this conclusion. The first is the claim that emotion figures in empathy only as the subject matter of the empathizer’s imaginative effort. A number of contemporary authors adopt this view at least intermittently. Jonathan Deigh, for one, suggests that empathy involves (for example) “imagining the feelings of frustration and anger, say, that he [the empathizer] would feel,” and explicitly clarifies that this state of imagining is to be understood as a cognitive state rather than an affective one.9 When we imagine a feeling, we picture it, but we do not have it. The other version of this conclusion instead has it that not emotion, but rather something emotion-like, is part of the empathizer’s imaginative effort. Kendall Walton argues that mental simulation (of which empathy is a sub-species) involves not imagining a feeling, but rather having a feeling-in-imagination: when I imagine going on a spelunking expedition, for example, “I might find myself, in my imagination, feeling strangely confident of my ability to cope, or being resigned to my fate.”10

---


confidence here should be described as a “quasi-emotion”, or a “make-believe” or “pretend” emotion.\textsuperscript{11} Because the emotion is strictly imaginary, on this view, there is no mystery as to why we do not act on it. Walton allows that mental simulation may be an affective experience. It can be unpleasant, or fun, or boring to imagine being in a certain position. But those latter, real feelings will be about the imaginative act, rather than what is imagined.

The two ways of describing emotion’s role in empathy exemplified by Deigh’s and Walton’s accounts will appeal to those who believe that emotion is inevitably conditioned on the belief that its object really exists.\textsuperscript{12} That thought has considerable staying power. It continues to fuel work on the “paradox of fiction,” which is a problem closely related to the one we are dealing with here. Certainly, our belief in the reality of an emotion’s object does seem to be required for the persistence of some emotions. If the grieving mother finds out that her son is in fact alive, she won’t carry on feeling sad just as before. It seems natural to say she now has nothing (or at least not the same things) to be sad about, and if her sadness persists, it must be because she doesn’t \textit{really} believe the good news. However, accepting the claim that that the empathizer \textit{qua} empathizer is not really in an emotional state comes at a high cost when it comes to capturing empathy’s phenomenology.

The fact is that empathetic experience does \textit{seem} to involve actual feeling. Adopting the other’s position, one registers the tragedy of their plight with a sinking sense of horror, or recognizes their victory with a thrill of joy. When I empathize, it doesn’t feel like I am engaged in a purely cognitive exercise. Nor does it feel like I am only pretending to feel some emotion, such that any real emotions I feel are not about what is imagined. It is hard to say what is involved in pretending to

\textsuperscript{11} Walton (2003): 261, 279.

\textsuperscript{12} For the classic articulation of this intuition, see Radford and Weston (1975).
experience some emotion, aside from putting on certain outward displays, but it is in any case not easy to reconcile the visceral, intense character of much empathetic experience with the thought that our apparent empathetic fear, joy and so on is only pretended.

Instead of denying that empathy involves emotion on the part of the empathizer, we ought to recognize, instead, that the fact that many emotions are conditioned upon belief does not entail that all of them are. Some emotions will fade away once we discover that their objects aren’t real, but others, including some that are quite powerful, will not. We can distinguish between “belief-oriented” emotions centered around what we believe to be the case, or at least what seems to us to be the case, and “thought-oriented” emotions that are oriented toward scenarios or objects that we are thinking about, rather than believing in. It is possible for me to be terrified at the thought of going deep sea diving, or thrilled at the thought of humans discovering a new Earth, even if neither of these scenarios is currently realized or even a real possibility.\textsuperscript{13} It is important to be clear that this view does not entail that when I experience these terrors or thrills, I am afraid of or excited about a particular mental object, a thought. I can be afraid of a thought if I believe that entertaining it will be detrimental to my health, but this that is not the sort of attitude that is relevant, here. Being terrified “at the thought of” going deep sea diving means being terrified about the content of the thought, namely going deep sea diving, despite knowing that my deep sea diving is not currently underway.

Thought-oriented emotions are by no means rare or exotic. They don’t have the same relation to action as their belief-directed counterparts do. Because pride at the thought of winning an award has a non-realized condition for an intentional object, it will not (for instance) lead us to crow about our

\textsuperscript{13} This kind of proposal has been defended by numerous “thought theorists,” including Lamarque (1981), Carroll (1990), and Moran (1994b).
achievement. But thought-oriented emotions still are a major part of our emotional lives. Thought-directed emotions about my own possible future states or actions, for instance, play an important role in shaping our planning and projects. If I find myself feeling queasy at the thought of working on Wall Street, for example, I may well take steps to avoid that possibility. And emotions about that which what no longer exists (which are also necessarily thought-directed) form an integral part of our sense of ourselves and our connections to others. Without them, I could not pity my younger self, our love those who are no longer living.

I suggest that we should understand empathetic emotion as a special category of thought-directed emotion, one specifically concerned with other people’s circumstances. The emotion that wells up in us as we contemplate the other’s situation from the inside is perfectly genuine. It is neither a make-believe emotion nor just a picture of a real emotion. Still, the background awareness that the circumstances we are contemplating belong to another, not to us, does influence how these emotions interact with our capacities to will, think, and feel. For one thing, when I maintain a firm grip on that background awareness, the pride that wells up in me the thought of success will not lead me to act in typically prideful ways. I will not crow or strut. Also, unlike thought-directed emotions about future possibilities for us, empathetic emotions do not primarily lead us to focus on our own planning and projects. So, for instance, fear at the thought of being left all alone that is empathetic, that is invited by another’s unwanted solitude, does not typically lead us to focus on what we can do to ensure our own robust social life. If we empathize properly, the primary effect of this fear will not be a retreat into memory of our past loneliness, either. Our overall response to our empathetic fear may include some hiving off of attention, such that some planning for our own future and reflection.

14 Piper (1991) is an illuminating discussion of how the vivid picturing of non-actual possibilities, which Piper calls “modal imagination,” matters to everyday life. She has in mind a phenomenon very much like the one I am discussing here, although she does specify that for most modal imagination our emotions will not be “fully engaged” (8).
on our own past is a marginal part of the empathetic experience, but what we are interested in when we empathize is principally the other’s situation and their experience of it. Therefore, when we empathize properly, we do not allow our emotion to orient our attention in ways that cause the other and the reality of their situation to recede from view.

I have offered an answer to the question of how empathetic emotions are special on behalf of Smith. It is not clear he would fully accept this answer. After all, he says that “the thought that they themselves [the empathizers] are not really the sufferers, continually intrudes itself upon them” as though this thought were an interruption that should result in the dulling of our empathetic feelings (TMS 22). According to my picture, by contrast, this thought should not be conceived of as playing an intrusive role in the empathetic process. It does not pull us back from a delusion that is part and parcel of empathy. On the contrary, an awareness of this fact shapes our experience of an emotion that owes none of its strength or durability to our confusing ourselves with the other.

3.3. The moment of judgment?

What comes after empathizer has done the imaginative work of placing himself in the mother’s situation, and felt some thought-oriented emotion well up in him in response to the picture of the situation he has created for himself? Smith describes this as the moment at which the empathizer forms a judgment of the other’s manifested emotion. He explains:

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. 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 16)
The empathizer compares the empathetic emotion she feels with emotion that she believes (on the basis of a report, or an expression, or some other grounds) the other is actually feeling. The two emotions will be in some degree of “concord.” If the concord is perfect, the empathizer will fully approve of the other’s emotional response to her situation. If they are not in concord, by contrast, the empathizer will disapprove of the other’s emotional response. So, in the case of the person empathizing with the mother: if he finds he can only summon in himself some mild discontentment at the thought of losing a son, whereas he is quite sure the mother is deeply distraught at her loss, he will disapprove of the mother’s emotion. This approval or disapproval with which an empathetic effort concludes, Smith is keen to emphasize, is not concerned with the strategic usefulness of the target’s emotion. Rather, it is concerned with whether the emotion is “right,” “accurate” “agreeable to truth and reality,” “precisely suited to its object” (TMS 20). Disapproving of the mother’s distress means concluding that her situation is not one worthy of such a dramatic response.

With the above passage, we come to what we can describe as Smith’s vision of empathy’s epistemic significance. Empathy matters to our epistemic position not because it is the sole route to knowledge of what others feel, but because it is the source of our beliefs about the propriety of others’ emotions. In this respect, Smith regards empathetic emotion’s function as continuous with that of our other “faculties”: “I judge of my sight by your sight, of your ear by my ear, of your

---

15 Smith speaks of the empathetic and the original emotion as in “perfect concord,” rather than as identical, because he takes it that the one will inevitably be weaker than the other, however closely they resemble each other in other respects. Given the picture of empathetic emotion I’ve articulated above, that the language of “concord” is suitable for a different reason: while the target of empathy may be experiencing a belief-directed emotion about her situation (as in the case with the mother, who is grieving what she believes to be the real loss of her son), the empathizer’s emotion can only match that original emotion insofar as it is its thought-directed analogue. Sadness at the thought of a loss is not precisely identical to sadness occasioned by belief in that loss, even though those two will obviously have much in common.

16 The final sentence of the passage actually seems to suggest an even tighter connection between the recognition of emotional concord and approval: a correspondence of sentiments just is approval. But I suspect the communicative intention behind this statement is much like Samuel Rogers’ “To know her was to love her”; Rogers’ point is that the knowing of “her” without the loving of her is impossible, not that when it comes to her case, loving and knowing are strictly identical. In that case, we can read this claim of Smith’s as establishing a bi-conditional: we approve of a sentiment if and only if we “adopt” it ourselves.
reason by my reason, of your resentment by my resentment, of your love by my love. I neither have, nor can have, any other way of judging about them" (TMS 19). Note the claim of exclusiveness here. Empathy matters to our epistemic position because it is the source of our beliefs about the propriety of others’ emotions. When it comes to accounting for empathy’s epistemic standing, our own emotions have a special role to play not because they are the only ones we can know with particular intimacy or immediacy, but because they are the only means we have of judging other’s attitudes.

Note that if Smith’s account of empathy’s epistemic significance were correct, we would have a clear explanation of how empathy matters to our lives as social, moral beings, one that does not hang on empathy’s ability to promote other-oriented concern. We could say that empathy is the means by which we are able to form judgments about others’ emotional perspectives, to determine whose outlooks are praiseworthy and deserving of emulation and whose are to be condemned, and that as such, it is an indispensable guide for conduct. However, Smith’s account of empathy’s epistemic significance is problematic in two respects.

4. Two problems for Smith’s proposal

Smith’s claim that empathy is the sole means by which judge the propriety’s of others’ emotion quickly came under critical attack from his philosophical contemporaries. In Thomas Reid’s 1780 lecture on The Theory of Moral Sentiments, he wrote:

[W]hen, [Smith] says, I judge your resentment by my resentment, your love by my love & so on, this is a way of speaking altogether new—this should be confined solely to our judging powers…to say that my resentment is the faculty by which I judge of your resentment is incommutable to all of our ideas of Resentment; we may as well say, I judge of my hunger by your hunger.17

---

17 From the “Lecture 100th,” March 1870. Transcription from Duncan and Baird (1977): 517.
In this section, I will suggest that Smith’s “way of speaking altogether new” is indeed flawed in two respects, though it is by no means true that Smith’s mistake is akin to the one I make when I posit that I judge your hunger by mine.

4.1. The first problem

The first problem with Smith’s proposal is that the particular good he associates with empathy is not, in our own experience, exclusive to it. Let us say that judging $p$ amounts to accepting the proposition $p$, where acceptance involves, roughly: taking $p$ on board for the purpose of answering $p$-related questions, being willing to affirm $p$ (except when strategic considerations make this unwise), and committing to acting in the light of $p$.\(^{18}\) In that case, when it comes to the judgment that some emotion is proper, there are indeed possible routes to that judgment that don’t pass via the judger’s empathetic experience. We can lean on the judgment of others in determining whether to accept propositions about emotions’ propriety. If my confidant says that our mutual acquaintance is reacting to his situation with unreasonable rage, and if I trust this confidant’s expertise in the domain of the outrageous, then I could accept her claim as true on the basis of her testimony, without needing to imaginatively place myself in our acquaintance’s position.

This is not just some hypothetical possibility, either: we are especially prone to basing our acceptance of propositions about emotions’ propriety on others’ testimony when we are mentally overburdened, or when we are still in the process of learning about the general conditions which render particular emotions appropriate. As we grow, we receive an endless stream of messages about the propriety of particular emotions under various contexts. It is not particularly strange that we

---

\(^{18}\) One might wish to exclude from the category of judgments cases of purely whimsical plumping for $p$. In that case, we can add that such acceptance must be based on evidence that $p$. For my part, I think that an acceptance of $p$ not based on evidence is best counted as a \textit{bad judgment}, rather than a \textit{non-judgment}, but the introduction of this additional requirement will not affect my argument in any case.
should defer to at least some of these messages, accepting the propositions they advance: the proposition that it is appropriate for a large person to be ashamed of her body, say, or the proposition that it is fine for a person to feel offended by another’s over-familiarity. Our acceptance of these propositions does not depend upon our ability to empathize with the shame and resentment of others who find themselves in these positions. We might accept them well in advance of our acquiring the imaginative ability to put our young selves in the position of people who are very different from us, whose preferences and concerns are largely beyond our ken. Therefore, we ought to reject Smith’s claim that empathy is the unique means by which we judge the propriety of emotions.

4.2. The second problem

The other problem with Smith’s proposal is that we certainly don’t inevitably judge all of our emotions to be as Smith puts its, “agreeable to truth and reality.” This seems like an obvious point. The evidence that we judge some of our emotions to be inappropriate will not be hard for any psychologically normal person to find. I could simply ask the reader to supply an example from her own life, and move on. However, in order to get at what is right in Smith’s picture of empathy’s epistemic significance, it will be helpful to a reflect a bit on when and how we can end up judging our emotions not to be all-things-considered appropriate. And to do this, we need to think a bit about what emotions are.

In some respects, emotions are a wildly diverse group of states. Some we experience as violent forces that seize control over us, while others seem comparatively easy to dismiss. Some are long lasting: an episode of grief may extend over a span of years. Others are over in seconds: fear may flicker in and out of existence when a threat quickly materializes and then vanishes. Some emotions
appear to be universal, whereas others are highly culturally specific. Impressed by this diversity, some philosophers have suggested that it is unprofitable to try to characterize emotions as a class. Amélie Rorty, for one, claims that it is a “notorious” truth that emotions do not form a natural kind, and that a “sensible person” would refuse try to investigate what makes an attitude an emotion.\footnote{Rorty (2004): 269.} However, I think it is worthwhile to seek the unity within this diversity, and I suggest we take our cue from Smith’s invocation of a resemblance between the functions of our “sight” and “ear” and the function of our emotions.\footnote{This may make me a less than sensible person in Rorty’s eyes, but I suspect that ship sailed long ago, anyhow.}

Emotions as a class are complex states comprising a number of different elements. For one thing, emotions typically involve familiar bodily symptoms. Anger, for instance, tends to manifest at least partly in a pounding heart, sweaty palms, and a feeling of heat. Disgust, by contrast, is associated with nausea and sometimes goosebumps. Emotions are also typically associated with tendencies to act and to express oneself in particular ways. In the case of anger, we speak harshly, stamp feet, refuse attempts at reconciliation. Importantly, those actions, expressions and bodily symptoms are not arbitrary. The many dispositions associated with a given emotion are unified by their connection to my recognition of certain features of the situation as inviting or demanding such irritated responses. Emotions focus our attention of some features of our situation, push others into the background, and present certain objects or events as calling out for particular kinds of treatment. So, in the case of anger, I attend to the intolerable smugness of the offender’s smirk, the outrageousness of their audacity, the sting of being disrespected. The relative triviality of the situation, and the good qualities of the offender that recommend forgiveness, recede into the background. That is to say, my anger presents the situation as laden with evaluative properties.
As the initial description above already suggests, I believe that we should think of emotions as being, at heart, ways of seeing in virtue of which things take on a particular evaluative sheen. In claiming that “we might as well say that I judge of your hunger by my hunger,” Reid might have meant to suggest that an emotion, like hunger, is simply a sensation or impulse that assails us, one that does not even have an intentional object. The view that emotions are simple sensations has had its defenders over the years. But while hunger comes in many forms, some of which may shade into emotional experience, the physical sensation of an empty belly is importantly different from emotion, because emotions share with other forms of perception, including the perceptions afforded by our “eye” and “ear,” the quality of being about some aspect or element of the world. We are mad about betrayals, sad about losses, thrilled about achievements. Importantly, this intentionality is manifest even in emotions’ somatic components. To appreciate this point, consider how the bodily feelings typical of an emotion are experienced by the one having the emotion. A lightheaded feeling that shows up outside of the context of any emotional experience will not register as intentional: the wooziness is not about anything, it just is. As a component of an emotional episode of joyous excitement, by contrast, a lightheaded feeling will typically be experienced not as a mere bodily sensation without intentionality, but rather as giddiness at the prospect of winning the prize, going on an adventure, or whatever the relevant enticing possibility is. And in another emotional context, a lightheaded feeling might instead be rightly understood as incipient panic at the prospect

---

21 The conception of emotions as modes of perception is one that has enjoyed robust philosophical popularity in the last thirty years or so. See e.g. Döring (2007), Tappolet (2016), and de Sousa (1990).

22 Most notably James (1884).

23 The idea that “emotional bodily feelings” can have intentional objects, and should sometimes be construed as “feelings towards,” is ably defended in Goldie (2000) and (2009). See also Ratcliffe (2008) for an argument that “many (but perhaps not all) of the phenomena referred to as ‘bodily feelings’ are not solely perceptions of our bodily states, distinct from the perception of our surroundings” (77).
of losing one’s livelihood, or reputation, or identity (and so on). The feelings embedded in these emotional contexts are bodily, but they nevertheless refer outward, to features of one’s situation.\textsuperscript{24}

Emotions are also world-directed in a particular way. Like sensory perceptions, emotions have a world-to-mind direction of fit. In Christine Tappolet’s helpful formulation, emotions, like other forms of perception, are “a kind of awareness of things and properties…a form of openness to the world.”\textsuperscript{25} Emotions are distinct from sensory perception in that emotional representations have an imperatival character: objects show up as to be avoided, or embraced, or destroyed.\textsuperscript{26} However, this fact does not prevent it from being the case that we experience emotions as disclosing or revealing aspects of reality. It’s just that the properties we perceive in emotional experience pertain to demands for action. So long as we are willing to allow that evaluative properties are in some sense real and not strictly subjective, then, we may conclude that an emotion, unlike either the sensation of hunger or a simple urge, can represent the world, and get things right.

\textit{Contra} Reid, the real problem for Smith’s view is not that emotions fail to represent, but rather that we sometimes judge our own emotions to be \textit{distorted} representations of reality.

Admittedly, this sort of judgment is not always a live option. To get a sense of the conditions that make such a judgment possible, it is important to distinguish between two different possible ways of

\begin{footnotesize}
\textsuperscript{24} In a number of psychological experiments in the 1960’s and 70’s, unknowing subjects’ somatic states were altered through the injection of drugs such as epinephrine. These studies have widely been regarded as furnishing evidence that individuals will interpret their bodily states as expressive of different emotions, depending upon available environmental cues (so, for example, an elevated heart rate might be interpreted by subjects either as a manifestation of fear or as a manifestation of romantic interest). See e.g. Schachter and Singer (1962), Schachter and Wheeler (1962), Dutton and Aron (1974). It should be noted that there are doubts about the reproducibility of these studies. See Marshall and Zimbardo (1979) for these worries.

\textsuperscript{25} Tappolet (2016): 30.

\textsuperscript{26} For this reason, some prefer to say that emotions have a mind-to-world-to-mind direction of fit: they represent both how things are and what is to be done. See e.g. Scarantino (2014): 177.
\end{footnotesize}
relating to an emotion. On the one hand, we can be entirely caught up in an emotion. In De Ira, Seneca describes anger as a “brief insanity…oblivious of decency, heedless of personal bonds, obstinate and intent on anything once started, closed to reasoning or advice,” a force that makes reason its “servant,” and “sweep[s] us on with a force of [its] own” (I.1.2, I.7.1). It is possible for anger to come over us as a kind of insanity, such that our rational capacities are entirely subordinated to the demands it places on us to lash out or to plot revenge. This is the kind of state that people reflecting on their own past episodes of anger frequently describe as “a runaway horse” a “fire,” a “flood,” an “eruption.”\(^\text{27}\) When we are in such a state, the possibility of either disapproving or approving of our angry representation of reality is not possible, precisely because there is no available reflective standpoint from which we can judge our anger’s representational accuracy.

That said, it is also possible for us to experience an emotion without being totally carried away by it. Emotion \textit{sometimes} functions as the highest authority when it comes to the determination of what to believe about the world, by effectively usurping the role of reflection, but (\textit{contra} Seneca) emotion does not always subsume our judgmental capacities. When we find ourselves able to recognize our emotion for what it is, namely one pattern of evaluative apprehension among others, we can either endorse it as genuinely revelatory of value or to judge it to be less than fully accurate. If we endorse the emotion as genuinely revelatory, then our emotional and judgmental capacities will, as it were, speak with one voice. We will be in a state of \textit{whole-hearted conviction}. Being in a state of whole-hearted conviction is quite different from being in the grip of an emotion, on a phenomenological level, because only the former involves recognition of one’s emotion as affording a perspective on one’s situation. From a behavioral point of view, though, those two states will look quite similar. In both

\(^{27}\) These metaphors are taken from a sample of men’s phenomenological descriptions of their own anger, in Thomas (2003): 167.
cases, our actions can be expected to align with our emotional apprehension of things as to be avoided, adored, destroyed, and so forth.

The alternative to a state of whole-hearted conviction is one in which we judge that our emotion is not wholly appropriate. If this latter judgment is not accompanied by the dissolution of the emotion in question, then our emotion is said to be *recalcitrant*. All sorts of emotions can stick around in spite of a judgment, made at a critical remove, that they are do not reflect the reality of the situation. So, for instance, I might continue to resent a friend for missing my party, even though I judge that she had a reasonable excuse for skipping out. Or, as in Hume’s classic example, I might fear falling over a cliff, despite judging that the iron cage I am encased in will absolutely prevent such a calamity, and that fear is therefore inappropriate (SBN 48). Emotionally recalcitrant states are uncomfortable to live in precisely because emotions don’t cease to make demands on us when we judge them to be distorted. We end up being torn between competing imperatives. Depending upon our the quality of our self-control, emotional recalcitrance may result in conduct rigidly disciplined by our judgment, or behavior very like that of the person in the grip of an emotion, or something between these two poles. The resemblance to sensory perception is particularly striking, here. We can judge an emotion to be inappropriate but still find ourselves seeing the world in its light, in much the same way that we can reject a sensory perception as illusory without yet being able to “shake it.” We can judge the deliverances of our own sight and hearing to be misleading (the stick in water isn’t really bent, there isn’t really a desert oasis shimmering in the distance), and yet still continue to be in some respects in their grip. In such cases, we can’t dissolve the appearance through the force of judgment, and we may still be inclined to behave as though it accurately reflects reality.
I have sketched out a picture of what emotion involves. Emotions are perceptions, and like their sensory kin, it is possible to experience them while judging the picture of the world they offer to be inaccurate. At long last, we can now bring these reflections about our psychology home to the case of empathy. If we do not, in fact, always judge our own emotions to be appropriate, it doesn’t seem like Smith can maintain that they, and they alone, are the standard by which we judge the propriety of other’s emotions, without making our practices of judging look incoherent, or at best seriously hypocritical. Nor will it do any good to revise the picture by saying that the only emotions of mine that I can properly use as a judgmental standard are those that I experience with whole-hearted conviction. The problem with such a response, of course, is that it makes the emotion itself look entirely judgmentally otiose. If I have to judge that my emotion is really reflective of reality before deploying it as a standard, I might as well simply apply whatever standards I use to ground that judgment directly to the other person’s emotion.

I’ve pressed on the tight connection Smith posits between empathy and judgment from two directions. First, I’ve noted that we can make judgments about the propriety of others’ emotions without recourse to empathy, because we can base our judgments on testimony about this propriety. Second, I’ve noted that I can have an empathetic emotion without judging it to accurately reflect reality. Those two observations together suggest that judgments of overall propriety are not the right place to look for what we are after, a special feature of our epistemic standing relative to our own inner lives that could via empathy be extended to the inner lives of others. Now, I'll turn to identifying what is right in Smith’s view of the connection between empathy and judgment. As it turns out, the problem with Smith’s account is not that there is no epistemic good unique to empathy, or that questions of propriety are totally the wrong place to look for an answer to the
question of what empathy can teach us. He has simply misidentified the mental state or activity that is key to empathy’s epistemic significance.

5. From judgment to humane understanding

Let’s return to the passage in which Smith first lays out the link between empathizing with an emotion and judging it to be proper. The first sentence of the passage is actually importantly different from what follows it, because it points to a connection between an empathetic match and the appearance of propriety, rather than between an empathetic match and a judgment of propriety: “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…” (my emphasis). Smith swiftly moves beyond talk of “appearances” of propriety, but I think that this is where we need to stay in order to get a proper handle on empathy’s distinct epistemic contribution.

I’ve already said that emotions should be understood as ways of seeing that present the world in a certain evaluative light. Let us revisit the case of anger to get clear on the connection between emotion and appearances of propriety.

Suppose I have just discovered that you passed along a secret of mine to your friends. I told you about my troubles in confidence, and now they seem to be public knowledge. I am angered by what you’ve done. My anger has all the hallmarks of a serious, though not overwhelming, incidence of that emotion. I feel hot and tense. My thoughts return again and again to your painful act of betrayal. I am on high alert for other offenses I can lay at your door. I reconsider whether we were ever friends: perhaps you have always been ready and willing to exchange my privacy for the thrill of
juicy gossip. Your open face now looks like a clever façade, and your cheerful greetings now seem calculated to artfully twist the knife you’ve stuck in my back. My anger presents a complex but decidedly filtered view of the situation as a whole. More particularly, anger presents the situation as demanding the behavioral and bodily responses characteristic of anger. As for behavior: it looks like I ought to dwell on your offenses, avoid you, perhaps lash out you (or maybe even plan some more elaborate revenge). This presentation of the situation also critically influences how I register my ongoing somatic responses. The feeling of surging adrenaline I might have otherwise taken to manifest high-spirited excitement or thrill feels to me, in these circumstances, like “boiling blood,” a bodily response to the outrageousness of the offense.

Because I am not entirely swept up in the emotion, it is open to me to conclude, on reflection, that anger is not all things considered appropriate. I might reach this judgment on any number of different grounds. Perhaps I think that anger can be a merited response to some situations, but not to this one (the offense is too trivial, say). Or perhaps I’ve been reading some Buddhist philosophy, and even though the thought of nirvana leaves me cold right now, I have concluded that anger is always a mistake, arising as it does from false views about the self and what “I” deserve.

One thing, however, is not available to me while I remain in a state of anger: I cannot both be angry at your betrayal and be totally mystified by my own anger. That might sound like a very bold claim. Once I clarify what I mean by “mystification,” however, it should become clear that the claim is no more bold than my earlier assertion that emotions are modes of evaluative perception. The claim about mystification is straightforwardly entailed by that analysis of emotion.

---

28 What I have to say here will apply only to the case of conscious emotions. For now, I set the possibility of unconscious emotion to one side.
When I am angered, I see something as outrageous. That way of seeing, as we’ve said, is the essence
of anger, the thing that unites its diverse behavioral, physiological and attention-directing features.
And when I see in this way, it is not an option for me to deny that my anger is tracking how things
appear to be, that it is reflecting the situation’s apparent evaluative features. When we appreciate that
our anger reflects the evaluative features the situation appears to have, the apparent presence of
those evaluative features (it looks like the situation is outrageous, like calls it out for revenge, and so
on) inevitably gives our emotion itself the appearance of propriety. Emotions are actually “proper”
when they accurately reflect a situation’s actual evaluative features, but they can appear to be proper
without actually being proper. The situation appears outrageous, so (of course), seeing it as
outrageous appears proper. We might use the language of intelligibility here: when we feel an
emotion, it cannot simultaneously be unintelligible to us, which is to say when we feel an emotion,
we cannot register it as failing to track the apparent qualities of the situation.

This is not to say that whenever we feel an emotion, that feeling is inevitably accompanied by a
reflexive awareness of its own intelligibility. We are not always aware of the intelligibility of our
emotions, and very small children are likely not capable of reflecting on their emotions at all. The
point is just that the first-person experience of an emotion constrains the reflective possibilities, for
us. I can simultaneously experience an emotion and be surprised by it. I can, for instance, feel that
that it is out of character for me, not something I can reconcile with my broader set of beliefs,
convictions and plans. But I cannot experience an emotion and simultaneously find it unintelligible.
And this is the sense it which it is right to say that my own emotions cannot mystify me.

Matters are quite different when it comes to a third-person perspective on an emotion. It is entirely
possible to regard others’ emotions as entirely unaccountable or wild, as failing to track any aspect of
an object’s evaluative appearance. Consider, for instance, Erika La Tour Eiffel, a women who has claimed to be madly in love with a series of large inanimate objects. It seems clear enough that she really did have romantic feelings for the Eiffel tower. She persisted in her courtship of the Parisian landmark even at the expense of being disowned by her family. On being banned from the tower, she lamented: “I don't even know how to articulate a heartbreak like that. It just wrecked me. It was this final blow, and I just had to withdraw.” It is entirely available to others to say that there is no sense to be found in her romantic love. Indeed, one of the reasons her case has attracted so much attention is that the public regards her feelings as not just inappropriate all-thing-considered, but totally unintelligible. There is absolutely nothing in the Eiffel tower that most of us can see as calling out for romantic devotion, as opposed to admiration, aesthetic appreciation, or (perhaps) fondness.

For Erika, however, the way that most of us see the Eiffel tower, as an emotionless, lifeless object that could never invite our erotic love, is not an option. Or at least, it is not an option so long as she persists in her love. If she were to judge that her love is all things considered misguided, she might try to dissolve the Tower’s appearance (to her) as calling out for her tender devotion. She could do this by working her way into a new perspective, in which only things that can smile, or think, or respond emotionally show up as possible objects of romantic love. Perhaps a rigorous course of cognitive-behavioral therapy could train her to experience architectural objects as unlovable. In that case, though, she would not actually remain in a state of love, either. She might still flush or feel her heart racing when she passes by the tower, but those somatic symptoms would be mere leftovers, no longer reflective of a genuine emotion.

---

What does this all mean for empathy? Well, the application is quite straightforward. We have said that our own emotions are inevitably intelligible to us, in the particular sense I have specified. We have also said when we empathize, we feel genuine emotion of a thought-directed variety. So, empathy involves feeling inevitably intelligible thought-directed emotions. Now, suppose I succeed in empathizing with someone. Let us stipulate, for the time being (1) I have an excellent knowledge of their situation, (2) I know what emotional response they are having to it, (3) I make no mistakes in imaginatively reconstructing their situation, (4) I know that I have made no such mistakes, (5) My empathetic emotional response is concordant with their original emotion (that is to say, my thought-directed feeling is the analogue of their belief-directed emotion), and (6) I recognize this concordance.

If all of the above conditions are satisfied, I will have fully appreciated the intelligibility of the other’s emotion first-hand. And in such a case I will not just be entitled to the conclusion that the other’s emotion is intelligible, I will be absolutely bound to accept it. This will be just as true whether or not I judge that my own empathetic emotion accurately reflects the evaluative facts about the situation in which I have imaginatively placed myself. So, for instance, in the case of the mother grieving the loss of her son: if the aforementioned stipulations hold, then when I empathize with the mother, I will see how the fact that her son’s future has been cut short appears to call out for bitter tears, and how the suddenness of his death looks like a shock that is impossible to accept. I might appreciate these appearances even if I myself reflectively endorse the stoic judgment that it is never really worth mourning any loss, no matter how severe.

Obviously, the conditions I’ve stipulated won’t all be fully met in all real-life cases of empathizing. Two sorts of misalignments are possible, each of which will have a different influence on the
epistemic outcome of my empathetic effort. On the one hand, there are the possible problems with the empathizer’s efforts. There will always be some imaginative slippage, and there will always be questions about whether I have missed some nuance of the other’s situation, or drawn the wrong conclusion about the character of their emotion. These possibilities will make it appropriate for me to modulate my confidence that the outlook I have managed to “get,” the one whose intelligibility is now plain to me, is precisely the outlook that the target of my empathy actually has. On the other hand, there is the possibility that I might fail to achieve full concordance with the empathizer’s emotion not because of some methodological imperfection on my end, but because my emotional sensibilities really are different from the target’s. In that case, the warranted conclusion will be that the other’s outlook is not fully intelligible. Of course, it is also possible for both of these sorts of misalignment to be present in various combinations. Still, grasping the intelligibility of another’s emotion is a substantial epistemic achievement, even if that grasp is partial.

I will call this grasping of intelligibility humane understanding of the other’s emotion. I believe understanding is the right way to characterize this epistemic achievement given its striking resemblance to the forms of understanding that pertain to other subject matters. Duncan Pritchard characterizes the understanding involved in an expert’s grasp of tidal movement, or atomic physics, or zebra biology, as a kind of “intellectual seeing.”30 The experts have understanding in these cases because they have an active and direct grasp of the relevant facts, one that underlies the experts’ abilities to explain the roles that various features of the phenomena play relative to each other and to the whole. In the case of empathy, the kind of seeing in question is not best characterized as an intellectual one, but the empathizer does actively and directly grasp the ways in which the other’s emotion.

30 Pritchard (2016). Some philosophical accounts of understanding distinguish between “objectual understanding,” where the target of understanding is some object, event, or body of information, and “propositional understanding,” where the target of understanding is a single proposition. See Kvanvig (2003): 195ff. Within this taxonomy, empathetic understanding is best construed as a form of objectual understanding.
situation seems to call out for the target’s emotional response. The empathizer doesn’t just know *that* the other’s emotion is intelligible. That, after all, is something that one could learn through testimony, without engaging in any kind of perspective taking. Rather, she directly apprehends the different aspects of the situation that look like they call out for anger as such. Typically, she will also appreciate, even if she can’t quite put it into words, how these aspects relate to each other. So, to take the case of empathy with anger, the empathizer will appreciate that there seems to have been a serious violation, that the object of anger looks responsible for that violation, that the violation the seriousness doesn’t seem to be mitigated by conditions that invite forgiveness, and so on.

It is important to note that humane understanding of another’s emotion is not merely a “weaker” version of a judgment that the other’s attitude is appropriate. We can judge others’ emotions to be all-things considered inappropriate and still deeply empathize with them, and (conversely) we can judge others’ emotions to be perfectly correct while finding that they leave us cold. For example, perhaps I am convinced by philosophical arguments to the effect that stoic indifference is the right attitude to have toward loss. I might still find, upon attempting to empathize with someone who exhibits such stoicism, that I cannot find my way into that cool outlook, no matter how much I strain my imaginative or emotional capacities. Approving judgment and humane understanding are often fellow travellers. Approving judgment often follows close on the heels of humane understanding, or is deeply intertwined with it, because we often (though not always) take our emotions to be a good guide to the actual evaluative properties of the situation. When we experience a party as delightful, we are generally very strongly inclined to conclude that is really *is*, unless we have some special reason to be unusually suspicious of our sense of joy. That said though, we should bear in mind that humane understanding and approval can come rather dramatically apart.
It is also worth observing, at this point, that in *directly* securing humane understanding of others’ emotions, empathy also *indirectly* secures humane understanding of others’ choices and actions, insofar as they are motivated by the agent’s emotional outlook.\(^{31}\) An otherwise baffling choice to undertake a grueling and dangerous solo kayak crossing of the Atlantic at age seventy will become intelligible to me if I can appreciate how the prospect of being “a little gray man” with a predictable and easy life might look hideously boring.\(^{32}\) I might still think that endeavor is all-things-considered foolhardy, but it will no longer be mystifying in the same way.

I have articulated the basic claim that empathy ensures humane understanding of another’s emotions. One of our stated desiderata for an account of empathy’s significance is that it conform with our own experience of our emotional and imaginative lives. Therefore, it is important to address a challenge to this claim that arises from an ordinary feature of our emotional lives, one we might describe as the apparent *inarticulacy* of some of our emotional experience.

Here is an experience whose general outline might strike readers as familiar. I have arranged to have dinner with an old friend at a favorite restaurant. I had made these plans enthusiastically. This friend and I go way back, and we haven’t seen each other in ages. But now, it strikes me that I am not wholeheartedly looking forward to the dinner. Instead, I feel on edge, in the grip of a creeping anxiety. I ask myself: What is souring my anticipation of the dinner? I consider that perhaps I am just nervous that we won’t be able to easily re-establish our old rapport, but that rings rather hollow—no, we’ve never had problems along those lines. I call up my sister to talk it over. She points out that while my friend has many wonderful qualities, she also has a bit of an irritating

\(^{31}\) Thanks to Quassim Cassam for pressing me on this point.

\(^{32}\) I am thinking here of the case of the “wild” adventurer Aleksander Doba, profiled in Weil (2018).
competitive streak. Perhaps what I am really dreading is the subtle but persistent one-upsmanship my friend tends to go in for. I mull over her suggestion, and decide that it hits the mail on the head. I am dreading listening to my friend steer the conversation back around to the specialness of her own accomplishments.

One might think that this kind of case gives the lie to my claim that we cannot both experience an emotion and find it unintelligible. My sister seemed to be more able to make sense of my emotion, to understand what my dread was about, than I was myself. If I couldn’t even initially say what exactly was bothering me, then why would it not be open to me to conclude that my emotion was unintelligible? And this isn’t an isolated, anomalous case, either. We often sort through our emotions with others, most particularly our friends and confidants, and rely on them to help us pinpoint what is bothering us about some exchange, or what exactly it is about some person that we find so appealing, and so on.

In response to this challenge, I suggest that the process of reflection in question here should not be understood as one in which I and my sister (or friend, or whoever) examine some object, my emotion, which is itself static and unchanged by our investigative work. It is not as though my anxiety in fact always had an equally determinate object. Rather, we should think of this process as one in which my emotion itself changes, in that it becomes more articulate. Talbot Brewer helpfully described this process of elaboration: “What do I do when a new but still quite indefinite feeling wells up in me? Well, sometimes nothing at all. But quite often I make it my project to determine what exactly is bothering me...The task is not to identify one’s mental state but to try to get clear one what one sees when one looks through it.”

---

As this observation of Brewer’s suggests, the analogy with perception is helpful here, too. I might initially have quite a vague, fuzzy visual perception of some kind of figure. But, by focusing on the form and asking myself questions about my visual experience (could it be a yeti? Is it perhaps a very large man?), I can refine my vision. My way of seeing changes as I look harder. In one sense, it is correct to say that I was always seeing the same thing. Throughout this experience, my perception was caused by the same object (a bear, say). But in another sense, my perception has shifted considerably. Initially, I had a visual experience of an indeterminate figure, but now my visual experience is of a large animal.

Likewise, we can grant that throughout my process of reflection, my anxiety had the same cause (namely, my friend’s tiresome competitiveness), and that I was not always aware of this cause. Initially, though, before talking with my sister, my anxiety was a vague kind of perception. The prospect of the dinner looked unappealing, but no one aspect of it was showing up for me as particularly inviting or demanding trepidation. That is not to concede that this vague perception was unintelligible to me. The prospect of dinner did look a certain way: it had the general appearance of being irksome, and insofar as it had this appearance, we can say that my negative emotional response to it would have been intelligible to me. But in talking with my sister, in (as it were) looking harder at the prospect of the dinner, my emotional perception acquired a more determinate object. I came to see my friend’s competitiveness as the dinner’s particular dreadful feature. My feeling is now about the tendency to competitive one-upsmanship, it is a perception of that tendency in particular, in a way it wasn’t before. And now, what is intelligible to me is my particular emotional perception of that competitive streak as irritating, rather than as (say) a lovable quirk.
I have proposed that we cannot regard our own emotional experience as unintelligible. I have argued that one consequence of this fact is that we cannot both empathize with another’s emotional experience and regard that other’s emotional experience as unintelligible. But this argument does not yet constitute a complete defense of the epistemic element of the humane understanding thesis. According to that thesis, empathy is the source of humane understanding. A complete defense of the thesis will therefore show that empathy is not just one means among others by which we secure humane understanding. Other ways of engaging with our fellows’ emotional experience may secure important epistemic goods of other sorts, but not this particular good.

To make the case that empathy’s characteristic epistemic good is unique to it, I will review and evaluate two different arguments that purport to show that empathy does not provide any special access to the intelligibility of others’ attitudes. The first, from Shaun Gallagher and Daniel Hutto, aims to show that the perspective-taking characteristic of empathy does not make a special contribution to our knowledge or understanding of how others’ attitudes “make sense.” Gallagher and Hutto instead maintain that others’ attitudes come to make sense to us through an alternative means, namely our mastery of narratives. The second argument, from Jane Heal, does embrace the claim that perspective taking is essential to the appreciation of the “intelligible sense or point” of other’s attitudes, but denies that the emotional engagement characteristic of empathy contributes to this appreciation. Gallagher, Hutto, and Heal are not necessarily targeting the precisely the connection between empathy and intelligibility that I defending. Still, the terms they use—“making sense,” an “intelligible sense or point”—do suggest they are aiming at something in its vicinity, and

---


evaluating their arguments should help to further clarify what humane understanding is, and what it is not.

6. The Narrativist Challenge

To place the narrativist challenge of Gallagher and Hutto in its proper context, it will be helpful to say something about the views they take themselves to be targeting. I am not the first person to suggest a connection between a form of perspective taking and the apprehension of some normative quality belonging to the target’s attitudes. On the contrary, the idea that such a connection exists has a rich past, particularly in the domain of the philosophy of history.

6.1. Some kin of the humane understanding thesis

In his 1949 book *The Idea of History*, R.G. Collingwood proposed that when it comes to the investigation of minds (others’ minds, most centrally, but also our own past mental life) the right approach is not a “scientific” one, but rather a “historical” one. Collingwood explains that the historian’s task is not just, or not even primarily, to develop an accurate record of the causal path that led to momentous decisions or events. Rather, the historian must discover the thoughts “within” historical events. The only way to discover these thoughts is through “re-thinking them in his own mind.” And provided that one is aware of what one is doing, Collingwood maintains, this re-thinking constitutes understanding: understanding just is the re-enactment of past thought. In re-enacting a thought, we appreciate the reasons that make an agent’s actions or choices “intelligible.” For Collingwood, an action is intelligible for one just in case one can think of it as a solution to a

---

38 Collingwood (2005): 283. Note the use of perceptual language.
problem one confronts. To appreciate the intelligibility of an emperor’s choice, for instance, one must “envision” the emperor’s situation as the emperor did, “see” how the situation might be addressed, and recognize that the solution one sees is also the solution the emperor in fact pursued. This activity, Collingwood emphasizes, is only concerned with the rational elements of the mind: thoughts, conceptions, and wills, rather than impulses, “feelings,” or appetites.39

Collingwood insists that the historical method is not only the sole means by which we come to appreciate the intelligibility of others’ thoughts and actions, but also the sole means by which we discover the contents of their thought. As he puts it, this is how “we discover the thought of the stranger who crosses the street,” and also, “it is only by historical thinking that I can discover…what I thought five minutes ago.”40 One difference between my position and Collingwood’s is that I do not insist that our knowledge of the contents of others’ mental states must arise through perspective taking, empathetic or otherwise. I am happy to join Adam Smith in his pluralism about the possible sources of this knowledge. That said, the link Collingwood posits between perspective taking and the appreciation of intelligibility is very much like the one I have been trying to establish between empathy and humane understanding.

Collingwood’s proposal bears resemblance to the “hermeneutical” approach to history favored by Wilhelm Dilthey, who argued that the correct practice of historical inquiry is the method of verstehen (understanding). This method involves the historian’s first-personal reconstruction of the target’s intentions.41 And both Collingwood and Dilthey’s proposals are echoed, to varying degrees,

41 Dilthey (1988); see also Outhwaite (1975).
in the more recent work of Karsten Stueber, Jane Heal, and Richard Moran.\textsuperscript{42} For ease of reference, I will borrow Moran’s term, and refer to this small group of philosophers as “interpretation theorists.”\textsuperscript{43} For Moran, an appreciation of the “rationality” of another’s intentional states involves thinking about the subject matter of the other’s intentional mental states. We think alongside the other, and thereby come to appreciate the respects in which the other’s thoughts are reasonable. Stueber similarly suggests that “re-enactive” empathy, which involves entertaining your beliefs and desires and deliberating from your perspective about what to do, is absolutely essential for understanding thoughts as reasons.\textsuperscript{44} Heal contends that “co-cognition” is necessary for apprehending another’s course of thinking as “the upshot of the exercise of cognitive skills,” such that the links between the other’s mental states show up as “rational or intelligible.”\textsuperscript{45}

These recent accounts, the admitted heirs of Collingwood’s theory, are more ambitious than mine in two respects.\textsuperscript{46} First, they inherit from Collingwood the ambition of accounting for our knowledge of the contents of other’s intentional mental states. Co-cognition, re-enactive empathy, or imaginative perspective taking is cast as the source of all or at least most of this knowledge.\textsuperscript{47} My account aims only to establish that the appreciation of a particular normative feature of others’ attitudes, their intelligibility, is tightly bound up with empathy. I will not try to defend the further

\textsuperscript{42} See Heal (2003); Stueber (2006); Moran (1994a).

\textsuperscript{43} Moran introduces the term “interpretation theory” in Moran (1994a). I include my own theory in this category, despite the differences I highlight below.

\textsuperscript{44} Stueber (2006).

\textsuperscript{45} Heal (2003): 78.

\textsuperscript{46} For explicit acknowledgements of the line of inheritance, see Stueber (2006): 14-21, 152ff and Heal (2003): 29, 77.

claim that the mere knowledge of others’ intentional states usually or always entails this kind of appreciation. Second, my account has a narrower subject matter than these others do. I am only concerned with emotionally engaged perspective taking, whereas these other accounts are also (or, in the case of Heal, exclusively) concerned with “cool” perspective-taking that does not involve feeling on the part of the one engaged in this imaginative activity. Another point of difference is that Stueber and Heal’s accounts tend to treat the property of being intelligible as either identical to the property of being rational, or as a weakened, less demanding form of that latter property. On my account, rationality and intelligibility can come rather more dramatically apart.

Despite these differences, my proposal does share with these other accounts the basic thought that perspective-taking is essential for appreciating a normative feature of others’ attitudes, one that has something to do with how those attitudes succeed in (or fail to) track the apparent or real features of the other’s situation. And it is this general idea that perspective taking is essential for what we might call “making sense” that Gallagher and Hutto are concerned to refute.

6.2. Narratives as the key to intelligibility?

Hutto and Gallagher endorse a “narrativist” approach to our acquaintance with other minds, and argue that our understanding of how both our own and others’ attitudes make sense is achieved not through imaginative perspective taking but rather through our deep familiarity and practical “skilled competence” with stories (such as fairy tales) that help children to grasp the way that actions,  

48 Stueber holds that we reveal actions to be “intelligible “ by revealing them “to be or to approximate to being, as they rationally ought to be” (2006): 49. Heal holds that the very same feature, “whether we can see what went on as the upshot of the exercise of cognitive skills, makes a link between intentional states “rational or intelligible” (2003): 78. Moran does not use the language of intelligibility.

49 I will use “making sense” as an umbrella term, one that covers both the “rationality or intelligibility” that interests Collingwood and his inheritors, and the more particular sense of intelligibility that interests me.
attitudes, beliefs and desires ought to fit together. So, in their view, we learn from Red Riding Hood that it will be proper for helpful person to want to being food to someone else, and to act on this desire when no greater concerns prevent her from so acting. We might also learn that it will make sense for a malicious creature to deceive others about his intentions. They explain:

Typically, [children] are provided with running commentaries on stories that teach them not only which actions are suited to particular situations but also which reasons for acting are acceptable and which are not. It is by absorbing such standards that we first learn how to judge an action’s appropriateness (though, of course, in time such standards are sometimes questioned and overturned). Quite generally, stories – real or fictional – teach us what others can expect from us, but just as importantly, what we can expect from others in certain situations. This is not just coming to know what others ought to (and thus are likely to) do, but what they ought to (and thus are likely to) think and feel, as indexed to the sort of people they are. (2008): 31

When we meet with surprising behavior, Hutto and Gallagher contend, our means of making sense of it will be to try to understand it in terms of one or more internalized narratives. And our grasp on these narratives effectively renders “simulation” (empathetic or otherwise) redundant: “Crucially, coming to appreciate the other’s story – to see why they are doing what they are doing – does not require a capacity for mentalizing inferences or simulations. Our understanding of others is ordinarily not based on attempts to get into their heads; typically we do not need to access a “landscape of consciousness” since we already have access to a “landscape of action” which is constituted by their embodied actions and the rich worldly contexts within which they act.”

Gallagher and Hutto roundly reject the proposal that any important aspect of our grip on others’ inner lives is derived from any kind of exploitation of a form of understanding native to the first-person perspective. Our understanding of how our own intentions, thoughts, and feelings make sense is a product of our absorption of narratives in just the same way that our understanding of

---


51 Gallagher and Hutto (2008): 34.
other’s intentions, thoughts, and feelings is.52

The attraction of this view is twofold. First, it is undeniably true that we learn a great deal about others’ inner lives, and also about the propriety of various thoughts, attitudes, and behaviors, through narratives. It is widely agreed that stories are a central means by which values are instilled and concepts are acquired.53 Second, an apparent special advantage of the narrative account is that it can easily explain how the intentions, thoughts, and feelings of a wide variety of characters, including characters very different from us, do “make sense” to us.54 To find intelligible the scheming of some cruel fictional character, it is claimed, we really just have to be intimately familiar with the relevant narrative trope: we know and appreciate, based on our familiarity with the relevant narratives, that it is fitting for evil characters to see things this way, to desire these outcomes.55 No perspective-taking mental acrobatics are required. It makes perfect sense to us that monstrous creatures like the giant of Jack and the Beanstalk fame have terrible wishes to tear children apart, even if our imaginations don’t stretch to seeing grinding Jack’s bones to make (our) bread as a worthwhile activity. Prima facie, theories that treat perspective taking as essential to uncovering the intelligibility of others’ attitudes will have a harder time accounting for this flexibility in our understanding.

52 Gallagher and Hutto are not the first to propose that our understanding of both ourselves and others is a matter of narrative competence. In the early 20th century, Wilhelm Schapp argued that both our understanding of ourselves and others and, more dramatically, our very being, is constituted by narratives. Schapp (2012); see Zahavi (2014) for a helpful discussion of Schapp’s view.

53 Work on the significance of narrative for social understanding has up until recently been largely the province of anthropology. Bascom (1954) is a seminal work on this topic.

54 For this claim, see Hutto (2007): 61.

55 See Kieran (2003) for an elaboration of this claim similar to Hutto and Gallagher’s. Goldie (2000) also develops a similar narrativist account. He claims that “In understanding and interpretation, we seek to locate the person’s episodes of thought and feeling, which go to form part of his emotional experience, in the overall narrative which makes best sense of this part of his life” (69). For Goldie, this work of narrative location need not involve first-personal perspective-taking. He clarifies that “Although the episodes which form part of the narrative structure are clearly emotional episodes, understanding of those episodes by the interpreter is nevertheless achievable third-personally” (189).
I will return to the question of how to account for our understanding of characters like the giant below. First, though, I want to highlight a critical tension that threatens the coherence of the narrativist account. For Gallagher and Hutto, it is critically important that the narratives in question have been suitably “internalized.” Why? Well, they want to avoid the possibility that one could “make sense” of others’ attitudes through the rote application of narrative “rules” that actually meant nothing to one. That consequence would threaten the distinction between their view and an implausible kind of “theory theory” according to which all there is to be understood about another person’s perspective could be adequately grasped by an alien being for whom familiar elements of human psychology are nothing but moving pieces in a theoretical system.56 According to the narrativist approach, there is supposed to be a difference in understanding between those who acquire practical mastery of narratives and those who merely memorize rules. Autistic people, Hutto contends, have only a rote grasp of folk psychological rules. They do not “internalize” the rules, and consequently they “never quite achieve the understanding of others that is the norm for most people.”57

The question I have for Hutto and Gallagher is what, exactly, internalization consists in. They sometimes use the language of “skill” or “know-how” to describe the difference between a person with rote knowledge of rules and the person who is genuinely able to appreciate how other’s

56 Hutto and Gallagher agree with the interpretation theorists that a theory-theoretic model of our apprehension of other’s mental states cannot account for all of our understanding of other’s mental states. According to thoroughgoing theory-theorists, our grasp of others’ mental states is secured purely through the exercise of theoretical reasoning, in which an observer applies general laws about connections between mental states, behaviors, and sensory input in order to predict and explain the others’ behaviors, choices, and mental states. For versions of theory-theory that treat the general laws as innate, see e.g. Carruthers (2013) and Scholl and Leslie (1999). For a version of theory-theory that treats the general laws as acquired through observation, see Gopnik and Meltzoff (1997). Ravenscroft (2016) offers a helpful overview.

attitudes make sense. But if skill or know-how in this case were to just amount to the ability to correctly identify others’ mental states, and to accurately deploy the labels of “reason” or “ought” to characterize those mental states and their contexts, it would not be clear why these abilities are not in principle something an autistic person (or, indeed, an intelligent alien) could perfectly master through sufficiently extensive rote learning. The acquisition of this kind of know-how therefore cannot constitute the difference between really absorbing narratives and exploiting them in the kind of rote, mechanical way Hutto and Gallagher characterize as the autistic person’s modus operandi. So, an alternative characterization of “internalization” is required—either an alternative description of the “skill” involved, or a characterization that replaces the reference to “skill” with something else entirely.

What else could the “internalization” that is by hypothesis not available to autistic people consist in, then? Here is what seems like a plausible answer. It is generally agreed that autistic people’s perspective-taking abilities are deficient; indeed, this lack of imaginative ability is considered a central feature of autism. Perhaps autistic people’s inability to internalize narratives stems from that imaginative deficiency. I propose that an adequate characterization of the internalization of narratives will refer to one’s ability to actually imaginatively see things in the way that narratives suggest. The appeal of this suggestion is (happily for my purposes) relatively easy to bring out in the case of emotional perceptions.

The critical difference between (a) merely memorizing Red Riding Hood’s story, and the related

---


59 See Baron-Cohen (2000) for an extensive review of the experimental evidence that autistic people’s perspective-taking abilities are deficient.
facts about the suitability of responses reported to us by adults, and (b) actually getting how Red Riding Hood’s terror makes sense, is most naturally described as the difference between merely committing the story to memory and actually finding ourselves able to envision “Grandma’s” big, sharp teeth as something scary, from Red Riding Hood’s point of view. If Grandma’s lupine characteristics just strike me as hilarious, no matter how hard I try to find something frightening in them, it seems intuitively correct to say there will be an aspect of the Red Riding Hood narrative, namely the heroine’s terror, that won’t strike me as sensible. The narrative will still be alien to me in a certain respect, even if I am able to exploit it skillfully in the service of identifying features that society has deemed terrifying.

Hutto and Gallagher claim that we can use narratives to appreciate the intelligibility of other’s attitudes only insofar as they are internalized. They further claim that autistic people do not internalize narratives, even though they may be able to exploit them through rote memorization and application. I’ve suggested that we can handily make sense of both of these claims by allowing that perspective taking is itself indispensible to the internalization of narratives.

One might now wish to point again to the case of the giant who wants to grind Jack’s bones to make his bread. Isn’t it true that the giant’s desire is perfectly intelligible to us, even though we cannot bring ourselves to imaginatively picture grinding up children as a worthwhile pursuit? Two things should be said about the case of the giant, and others like it. First, this monstrous wish makes some sense to us because we are able to get into it on a very general level. I grant that most of us can’t see bone-grinding in particular as a good thing, but we can see taking steps to alleviate hunger as desirable (the Giant’s wish in this case thus makes more sense to us than would the desire to, say, grind Jack’s bones even though the Giant will hate every minute of the task and get nothing out of
The apprehension of intelligibility is not an all or nothing matter. And second, there is a difference between narratives’ being familiar and their actually making sense, even though familiarity can sometimes be mistaken for intelligibility. Our early exposure to stories of monstrousness keeps us from being surprised when we encounter descriptions of such beastliness, but a lack of surprise should not be equated with sense-making. Even utterly bizarre behaviors can start to become banal if we are exposed to them long enough. So, the giant’s desire is fully familiar and predictable to those of us who have heard stories like it enough, but it is only partially intelligible.

So far, I’ve been making the case that narrative competence is not a substitute for perspective taking, because it actually requires perspective taking. One possible argumentative route for Hutto and Gallagher would be to concede that perspective-taking is necessary for establishing the kind of narrative competence that matters to sense-making, but hold that it is like ladder to be kicked away. Once our grip on narratives is in place, we won’t need to imaginatively shift our point of view in order to get all the understanding that anyone could hope or expect from us.

I’m also skeptical of this more concessive position. Undoubtedly, it is true that we often take shortcuts in thinking about the normative status of other people’s attitudes. This is a point made by Adam Smith himself, who observes that when we pass a stranger on the street “with all the marks of

---

60 This claim has an affinity with Zahavi’s criticism of Schapp’s narrativism (see footnote 54 above). Zahavi argues that “the narrative account has to be supplemented by an account that specifically targets the first-person character of our experiential life. In order to tell stories about one’s own experiences and actions, one must already be in possession of a first-person perspective. To that extent, experiential ownership is a prelinguistic presupposition for any narrative practice, rather than the outcome of active storytelling” (2014): 59.

61 At times, they seem to flirt with this position. At (2008): 32, most notably, they suggest that narrative competence “presupposes a wide range of emotive and interactive abilities. To appreciate such stories children must initially be capable, at least to some degree, of imaginative identification and of responding emotively, just as they do in basic social engagements.” This mention of “imaginative identification” suggests that they have in mind something like my proposal, but that would seem to run counter to their contention that understanding other minds does not involve exploiting a form of understanding particular to the first-person perspective. For more debate about exactly where the disagreement between interpretationist theories and narrativist theories lies, see Stueber (2012), Hutto (2012).
deepest affliction; and we are immediately told that he has just received the news of the death of his 
father,” we may approve of the stranger’s grief on the basis of the knowledge that if we properly 
entered into his perspective, “we should, without doubt, most sincerely sympathize with him” (TMS 17-18). It is not as though everyone’s attitudes show up as entirely bewildering to us unless and until we deeply imaginatively engage with their situations. For our own ease (and sanity), we may rely on internalized patterns in assessing others’ attitudes, a kind of shorthand based on our cultivated stock of narratives. That said, whether or not we recognize another’s attitude as fitting with narratives we’ve already internalized doesn’t settle the question of whether that attitude really makes sense to us, because we may find that the narratives we’ve embraced are actually deficient. Perhaps they cast as reasonable some attitudes that we find ourselves resisting when we actually try to imaginatively inhabit the other’s position, or as unreasonable attitudes that, when the particularities of a case are taken into account, we find we can empathetically work our way into. An attitude can make sense not because it conforms to standards or stories we’ve previously embraced, but because it turns out to be a way of viewing or otherwise relating to the world that “rings true” for us when we imaginatively enter into the other’s situation.62 And these kind of perspective-taking discoveries will lead us to introduce more subtlety into the narratives we employ.

In insisting upon an important role for perspective taking in the development of our understanding of how others’ attitudes make sense, I do not mean to deny that narratives help us to profitably deploy our capacity for feeling our way into other perspectives. Narratives can support and enhance our perspective-taking efforts by helping to fill out the details of a character’s situation for us. These details help to guide our imaginative efforts because they fill in gaps in our apprehension of the other’s situation that we might otherwise be inclined to erroneously “color in” with details borrowed

---

62 See Heal (2003): 20 for a similar point, discussed in section 7 below.
from our own experience, or (if we are more cautious) to simply leave blank. Simply being presented with the details of another’s experience in narrative form also tends to pull us into the activity of imaginative perspective taking. Stories of all forms (literary, historical, fanciful…) naturally excite our imagination: it is no accident that Adam Smith’s descriptions of our perspective taking so frequently focus on “the heroes of tragedy or romance,” (TMS 10), “the description in a journal of a siege, or of a sea voyage,” (TMS 30), and what “we read in history concerning the perfidy and cruelty of a Borgia or a Nero” (TMS 76). It might well be that without the exercise of our imaginative capacities that narratives naturally prompt and foster, our perspective-taking abilities would be seriously underdeveloped. We can fully embrace that idea without concluding that narrative practice renders perspective taking redundant, though.

7. The Co-cognitivist challenge

I will now turn to a challenge to the humane understanding thesis internal to the camp of interpretation theorists. This challenge, which comes from Jane Heal, takes no issue with the claim that a form of perspective taking is essential to the appreciation of some normative feature of others’ attitudes. Instead, the challenge is directed at the proposal that the emotional engagement characteristic of empathy contributes to the appreciation of this normative feature.

As mentioned above, Heal argues maintains that “co-cognition,” the “exercise of one’s cognitive powers on the content of the other’s…judgments,” is the sole means by which we are able to grasp the “rationality or intelligibility” of others’ attitudes.63 We can’t grasp the rationality or intelligibility of others’ attitudes by asking whether they conform to any particular set of rules, because rationality

63 (Heal 2003): 83. She also thinks it is the central means by which we are able to predict and explain others’ thoughts, but those latter two claims will not concern me here.
is (per Karsten Stueber’s helpful phrase) “essentially contextual.” In our own case, we can make our departures from any rigid set of rules of rationality “intelligible to ourselves.” Unless we wish to embrace the highly unattractive conclusion that the appropriate standard for rationality or intelligibility is very different in the case of others’ attitudes than it is in the case of our own, then, we should accept that a proper assessment of the rationality or intelligibility of others’ attitudes will involve not comparing their attitudes to some rule set. Rather, it will involve the following steps. In order to determine whether the other’s attitude $a$ is rational or intelligible, given that they believe (say), $x$, $y$, and $z$, we first assume the truth of the other’s relevant beliefs $x$, $y$, and $z$. We then ask ourselves: “Is it true that $a$?” If the answer is “Yes, $a$,” then (and only then) we may conclude that the other’s attitude is rational, or at least that “some intelligible sense or point can be seen in it.” In such a case, we will have effectively appreciated the rationality or intelligibility of the other’s attitude for ourselves.

Up to this point, Heal’s story parallels mine quite neatly. The fact that she calls this exercise “co-cognition” signals, however, that she has a different conception of what this exercise will look like as far as emotional states are concerned. Heal claims that the only attitudes to which the co-cognitive method of understanding will apply are “items with content,” a point upon which we agree. Itches, tingles, and other such sensations are in and of themselves non-intentional states (which I take it would in Heal’s terms make them “items without content”), and as such they are not the kind of states that could be rational or intelligible in and of themselves. But Heal takes it that limiting the scope of co-cognition to items with content spells trouble for the traditional picture of “what is involved in simulating a desire, emotion or intention,” according to which “simulation involves a

---


kind of faint version of what is simulated, conjured up by the imagination,” or instead “the thing itself [that is to say, the desire, emotion, or intention] but occurring in the relevant ‘mental system’ while it is being run offline.”\textsuperscript{66}

Now, the two particular accounts of empathetic perspective taking that Heal targets here are not strictly the same as mine, since I have proposed that empathy involves genuine feeling that is thought-directed. That said, my account would clearly be tarred by the same brush. Heal is suggesting that when it comes to appreciating the rationality or intelligibility of others’ mental states, the only simulation that matters is the entertainment or supposition of the mental states’ contents. Heal does not make Reid’s (apparent) mistake of casting desires, emotions or intentions as contentless, non-intentional states. She grants that desires, emotions and intentions do feature a contentful element, namely the judgments they contain (or, as she puts it, are “associated” with).\textsuperscript{67} However, she proposes that the intelligibility or rationality of these states can be fully appreciated through a simulation that concerns only the “associated judgments.” One of the essential features of empathy, as I’ve described it, is that involves rich imagination in contrast to supposition, and more particularly that it involves emotionally engaged imagination. On Heal’s view, by contrast, empathy in this sense bears no special connection to the appreciation of intelligibility.

Other philosophers who have tried to justify the intuitively appealing claim that empathy secures a special epistemic good for the empathizer have been troubled by the apparent inevitability of conclusions like Heal’s. Justin Steinberg, for one, argues that we must reject exclusivism (which he presents as the thesis that empathy is the exclusive source of some knowledge or understanding) on

\textsuperscript{66} Heal (2003): 81.

\textsuperscript{67} Heal (2003): 81.
the grounds that “The case for exclusivism turns on the plausibility of so-called phenomenological facts,” and “if there are phenomenal facts, their character is ineffable, their content defies specification; they can only be experienced, never described.”\textsuperscript{68} Like Heal, Steinberg seems to be working with a conception of emotion according to which emotions are decomposable into two elements.\textsuperscript{69} On the one hand, we have an evaluative judgment. And on the other hand, we have the equivalent of itches and tingles, phenomenal elements that make emotions \textit{emotional}, but that are themselves entirely without content.\textsuperscript{70} It is assumed that the non-judgmental elements of emotional experience can be prized away from the judgmental elements for the purposes of simulation, without thereby altering or distorting the judgmental element. Whether the judgment is embedded in an emotional experience or made in a coldly intellectual manner does not affect its content or its intelligibility. Heal’s reliance on this kind of picture comes through clearly in the following summary of her position with regard to emotional simulation:

Phenomena like depression, akrasia, overreaction and the like show that strength of motivation or feeling can get out of line with what is rationally licensed by the associated judgments, in being either too weak or too strong. Hence there is more to desire, emotion, and intention than such judgments– there are the characteristic causal links with purposive action and with a disposition to expressive behavior. But it does not follow from this that there is more to simulating a desire, emotion, or intention than entertaining the content of the associated value judgments. As far as rationalizing and making intelligible are concerned, it is the value judgments that do the work. Heal (2003): 81

In the face of these claims of Heal’s, why should we still think that there is indeed an indispensible role for emotional engagement when it comes to the appreciation of the intelligibility of others’ attitudes? To answer this question, I suggest that we return to the relation between emotional

---

\textsuperscript{68} Steinberg (2014): 53.

\textsuperscript{69} In contrast to Heal, Steinberg doesn’t discuss the case of desire or intention. From here forward, I will focus only on what these views have to say about emotional states.

\textsuperscript{70} See Matravers (2011) for a very similar point.
experience and intelligibility in the case of our own attitudes. If it can be shown that intelligibility is inextricably bound up with emotional experience in our own case, then we will have very strong grounds for the corresponding conclusion when it comes to (at least some of) others’ attitudes, too.\footnote{This argumentative strategy is one that Heal herself should accept, since she herself relies upon the claim that the appropriate standard for rationality or intelligibility should be the same in the case of others’ attitudes and in the case of our own.}

I will now argue that experiencing an emotion is insispensable for the appreciation of its intelligibility. We cannot appreciate the intelligibility of our emotions, we cannot properly grasp how they make sense, from a standpoint that is, as it were, not their own. I realize that that is a controversial claim. It is also hard one to assess, in the first person, occurrent emotion case, simply because we will inevitably be having the emotion we are meant to be assessing unemotionally. But to make the proposal more palatable, let us consider a case where the emotion in question is one we are currently considering, rather than currently experiencing.

Heal specifically notes that phenomena like “depression” are beyond the bounds of what matters to an appreciation of intelligibility. So let’s see what difference being in a depressed state can make to intelligibility in the sense that concerns me. Here we have a person who is attending a birthday party while in the grip of intense ennui. She is contemplating the possibility of feeling delighted by the party. Perhaps our bored, anhedonic subject acknowledges, intellectually, that joy should be the response for her to have. The smiling faces of friends, the upbeat music, the clink of champagne glasses: she knows that these are the kinds of things that normal people find delightful, that she would normally find delightful. But despite knowing that these are the things that invite joy, she now finds herself unable to find them delightful, to experience them as wonderful. Her heart does not
skip with happy excitement. Instead, everything looks dull, pathetic, alienating. And given that the party looks this way, the prospect of being delighted by it lacks a certain intelligibility for her.

Perhaps our partygoer approves, overall, the prospect of a joyful response to the party: she endorses the claim that the situation is a joyful one. But joy in this case would in a very real respect not make sense for or to her, because it would not be responsive to the evaluative qualities the features of her situation appear to her to have. The situation appears to merit a downturned mouth, a shiver of disdain, a miserable dwelling upon the ephemerality of happiness. This is a case in which what one approves as a rational response to a situation diverges from what one grasps as an intelligible response to the situation’s apparent evaluative qualities. For the intelligibility of a joyful response to be secured, she would have to change her way of seeing—which would mean changing her emotion.72

If I am right about this case, what does it mean for empathetic engagement? Suppose that I am empathizing with a person who is experiencing a party as a joyful occasion. She is full of cheer and delights in the music, the laughter, the dancing. In order for me to fully appreciate the intelligibility of her emotion, I will have to do more than merely coldly judge that the situation is a joyful one. That judgment, as we’ve seen, could be compatible with a failure to imaginatively perceive the party

72 The claim that what is intelligible to one can diverge from one’s evaluative judgments is also defended in Stocker and Hegeman (1996) and Johnston (2001). Johnston’s account is particularly persuasive. It does differ from mine in two respects. First, it focuses on affect, rather than emotions, although for him the category of “affect” extends to cover “something akin to sensing and sense-based imagination” which registers “determinate sensuous values” such as “the beautiful, the charming, and the erotic”: the resemblance to my treatment of emotion should be clear. And second, Johnston is concerned with the intelligibility of actions stemming from evaluative perception, rather than the intelligibility of the evaluative perceptions themselves. However, his observation about the connection between affect and intelligibility strongly resonates with the account I have been articulating, and it is simply too good not to quote: “[S]uppose I do bring myself, by sheer force of will, to munch on [a] Power Bar. I’d like to acquire even more gravitas, so I fancy putting on some more weight. To that extent, I believe it would be in some respect good to eat the bar. Yet my munching on the Power Bar can still lack a certain intelligibility even though I judge it good in some way. The thought ”Why on earth am I doing this?” can coexist with judging that there is something to be said for eating the Power Bar. For I can still fail to see anything appealing about the Power Bar” (2001): 190.
scene as anything other than a pathetic attempt to distract from the bitter truth that life is a tale told by an idiot, full of sound and fury, signifying nothing. Instead, I will have to work my way into actually experiencing a spark of joy at the thought of attending the party. If I am depressed, I will be unlikely to find this ability within myself, and so other people’s celebratory jubilation will appear unwarranted, even if I approve of it. On the other hand, I may be uniquely posited to effectively empathize with the glum partygoer, whose gloomy perception I do not endorse any more than she does, but whose outlook may nevertheless seem to perfectly reflect the apparent evaluative features of the situation.

It is interesting that Heal dismisses phenomena such as akrasia and depression as irrelevant when it comes to appreciating the normative status (the “rationality or intelligibility”) of emotions via perspective taking, because empathy seems to especially shine in cases where the target of our empathy is akratic or depressed. Most of us can imaginatively work our way into an akratic reluctance to get up in the morning, despite a resolute judgment that walking the dog is what is urgently called for: the cozy bed seems to call for us to snuggle down. Akrasia is a prime target for commiseration. It is the paradigm of an irrational condition, and yet there is something utterly emotionally intelligible about our akratic perception of the situation. A depressed outlook might be less universally accessible, and not so easy for one without firsthand experience of melancholia to work one’s way into. Still, some people undoubtedly can empathize with depressed states, and find the sense in the bleak view of the world the sufferer cannot help but by taken in by, despite her failure to endorse it. In maintaining that appreciating the “rationality or intelligibility” of another’s emotion is just a matter of co-cognizing, then, Heal passes over a kind of sense-making that seems eminently real. In the next chapter, I will show that the ability to secure humane understanding of outlooks like these is not just an interesting quirk, but rather a capacity of real moral significance.
8. Conclusion

The previous chapter explored the Humean way of developing the bridging picture of empathy’s moral and epistemic significance, and concluded that it was not viable. I have now introduced an alternative to the Humean picture, the humane understanding thesis: empathy is the source of a distinct epistemic good, humane understanding, which consists in the appreciation of the intelligibility of other’s emotional perceptions, and humane understanding is necessary for (and, indeed, partially constitutive of) fully virtuous relations with other people.

This chapter focused solely on the defense of the claim that empathy is the source of humane understanding. In order to defend that claim, I drew upon Adam Smith’s *The Theory of Moral Sentiments*. Smith contended that empathy is the unique means by which we judge of the propriety of others’ sentiments. I have accepted the broad Smithian thought that we need empathy in order to apprehend a normative feature of others’ emotions, but I have rejected Smith’s more specific claim that empathy is necessary and sufficient for judging whether others’ emotions are “agreeable to truth and reality.” Instead, I have proposed that the epistemic good that empathy uniquely affords is a kind of understanding. When we succeed in empathizing with another’s emotional outlook, we grasp firsthand how their emotion tracks the apparent evaluative features of their situation.

One of the desiderata for an account of empathy’s significance, as presented in Chapter One, is that an account of empathy should reveal the significance of each of empathy’s characteristic features. Empathy is imaginative and emotionally charged, and it involves getting at others’ experience from the inside. In the final sections of this chapter, I have addressed two challenges, both of which target the relevance of one of these features to our ability to grasp the “rationality” or “intelligibility” of
others’ outlooks. The first challenge questioned the need for perspective taking, while the second targeted the necessity of emotional involvement on the part of the empathizer. The latter challenge is, I think, reflective of a more general difficulty that has emerged over the course of the last few decades in the philosophy of emotion. Philosophers have been eager to acknowledge the intentionality of emotions, but this has in certain cases led to an over-assimilation of emotions to judgments. The possibility that emotions’ intelligibility can come apart from their rationality has consequently receded from view. The account of emotions I’ve offered here, which casts them as evaluative perceptions, helps to make clear how this divergence can occur.

Returning to the humane understanding thesis in particular, one might wonder how the thesis as I have elaborated it thus far relates to the bridging picture whose broad appeal I described in Chapter One. At the very most general level, it will be recalled, the bridging picture is the conjunction of two thoughts. First, there is the thought that our own thoughts and cares tend to occupy very different positions in our mental economy than do others’ attitudes and needs. And second, there is the thought that empathy allows us (sometimes, and imperfectly) to overcome this usual gap between our self-directed knowledge and/or concern and our usual or primitive other-directed knowledge and/or concern by exploiting something special about the first-person perspective. The proposal about humane understanding I have articulated in this chapter is not an alternative to the bridging picture. Rather, it is a version of it, albeit a very different one from Hume’s. What is special about and original to the first-person perspective, I have suggested, is that when we experience an emotion “from the inside,” we cannot coherently dismiss it as unintelligible. When the question of our own emotions’ intelligibility is raised, we will necessarily appreciate that our own emotions track the apparent evaluative features of our situation. Viewed from the outside, others’ emotions can look
entirely unintelligible, but when we imaginatively adopt another’s perspective in an emotionally engaged way, we cannot fail to appreciate their intelligibility.

Now, it may look like my account makes empathy out to be a thinner, less robust bridge between self and other than Hume and the Humeans wanted it to be. But at least it is not vulnerable to collapse, as accounts in the Humean line are. It does not require us to erase the differences between ourselves and others in order to work, nor does it make empathy’s efficacy as a bridge depend upon a mechanism on a par with the tendency for one’s anger at one person to bleed into one’s interactions with another. And, as I will show in the next chapter, the humane understanding we secure through empathy will influence our relations with others in several very morally important ways.
CHAPTER FOUR

The Moral Value of Humane Understanding

1. Introduction

The previous chapter was devoted to establishing that humane understanding is an epistemic good available uniquely through empathy. That chapter was partly aimed at the vindication of the intuition that empathy’s feeling nature matters to the quality of the knowledge (or, rather, understanding) that it affords, an intuition that Hume’s account of empathy’s significance did not seem to accommodate. Now I will turn to the vindication of the other guiding intuition that does not sit happily with Hume’s account, namely the thought that the way empathy matters morally has everything to do with the way it changes our understanding of others’ inner lives.

Earlier, I noted that Smith had the benefit of an obvious story to tell about the connection between empathy’s epistemic and moral significance. If empathy were necessary and sufficient for approving of others’ evaluative perceptions as correct, then it would obviously be hugely important in guiding our determinations as to which perspectives ought to be admired and affirmed. One might wonder whether, in shifting from talk of approving and judging to talk of appreciating the intelligibility of others’ emotions, we are also, regrettably, moving away from any such obvious explanation. There might be a lurking suspicion that humane understanding is, really, not worth much in the grand scheme of things. What could be morally important about appreciating the intelligibility of another’s emotional outlook, if that appreciation comes apart from all things considered approval or disapproval?
As it turns out, humane understanding is worth quite a bit. It is multifariously morally important. What I will offer in this chapter is not an exhaustive list of all the ways in which empathy might make a difference to our relations with others. Rather, I will explore what I take to be several particularly important and interesting aspects of its moral significance, ones that have by and large not yet received the attention they deserve. Some of these aspects concern empathy’s moral developmental significance, the ways that our pursuit and attainment of humane understanding shapes or distorts our relations with others over the long term. Others instead have more to do with what is gained or lost, morally speaking, in individual cases where one manages to empathize, or fails to. It will emerge that humane understanding is not always and everywhere an unqualified good that we have moral reason to pursue. Its goodness is highly dependent upon context, and I will describe some contexts in which the pursuit of or reliance upon humane understanding will be morally useless or even vicious, rather than virtuous. The fact that such contexts exist does not undermine the intuitively appealing claim that empathy is the great instrument of the moral good. It just means that empathy, like many other morally important and valuable phenomena, has its dark side.

In recent years, a couple of theorists working at the intersection of philosophy and psychology have defended the provocative thesis that empathy is very far from being a powerful force for the moral good. Jesse Prinz and Paul Bloom argue that empathy is something we would all things considered be morally better off without.¹ To prepare the way for my own claims about how empathy’s provision of humane understanding matters, I will begin by reviewing Prinz and Bloom’s arguments.

¹ They have pursued this argument in a number of places, including (in Bloom’s case) some non-academic venues. See Bloom (2014), Bloom and Zaki (2016), Bloom (2016), (2017a), (2017b) and Prinz (2011a), (2011b).
Prinz and Bloom launch attacks on what they take to be the two *prima facie* appealing reasons to think that empathy is essential to our moral lives. They contend that contrary to what we might have thought, empathy does not contribute anything special to our moral epistemology (our ability to know or understand the good, the right and so on). Nor, they contend, does it make any irreplaceable contribution to our moral motivation. And if empathy makes no indispensable contribution in either of these respects, they conclude, then empathy is not morally necessary. It doesn’t produce any morally good effects that we can’t come by through other means. Because the other means by which we can secure moral knowledge and motivation are less problematic than empathy, not being affected by the kinds of bias and shortsightedness that (it is argued) empathy inevitably involves, Prinz and Bloom conclude that empathy is worse than morally useless. We can do without it, and we should.

Reflection on the value of humane understanding will allow us to see why Prinz and Bloom are wrong in some critically important respects.\(^2\) My strategy for the rest of the chapter has two prongs. I will first make the case that there is something wrong with an assumption that is built into Prinz’s and Bloom’s arguments. They presuppose that empathy’s moral significance is a function of its consequences for our moral knowledge or motivation. They do not take seriously the possibility that there can be something morally good about succeeding in empathizing (which is to say, securing humane understanding) independent of its consequences in either of these domains. I argue that this is a mistake. For many of us, much of the time, receiving humane understanding from others is valuable in and of itself. It is so valuable to us that its absence can be a source of very real pain. Therefore, humane understanding is something we will often have moral reason to supply to others, and not just for its further moral motivational or moral epistemic consequences. I will also make the

\(^2\) A significant number of philosophers and psychologists have criticized Prinz and Bloom’s anti-empathy stances. See Masto (2012), Passos-Ferreira (2015), Persson and Savulescu (forthcoming), Borg (2014), and Song (2015).
case that humane understanding is non-instrumentally morally valuable as a necessary part of friendship.

Turning to the second prong of my argument, I will then show that Prinz and Bloom’s critical arguments miss out on or unduly minimize ways in which empathy is important even when it comes to the domains they do consider, those of moral epistemology and moral motivation. On the epistemic front, I will suggest that empathy is a means of expanding and deepening our sensitivity to moral features that we as human beings would be very hard pressed to do without. And on the motivational front, I will suggest that while empathy is not necessary for moral motivation, it can serve as an important check on some of our darker impulses.

2. Against empathy

Before delving into Prinz and Bloom’s criticisms, it is important to establish that the phenomenon they have in their sights is empathy in my sense of the word. This certainly seems to be the case for Bloom. Bloom describes empathy as “the act of coming to experience the world as you think someone else does,” and explicitly indicates that he is concerned with the same phenomenon that occupied the Scottish Enlightenment philosophers. Empathy in his sense implicates our emotion, and is also the kind of thing we have in mind when we talk about stepping into another’s shoes. Prinz’s definition of empathy does not explicitly make reference to looking out at the world from the other’s position. He invokes Hume in defining empathy: “[E]mpathy is a matter of feeling an emotion that we take another person to have. Following Hume, we can think of empathy as a kind of associative inference from observed or imagined expressions of emotion or external conditions

that are known from experience to bring emotions about.”  

It is unclear if Prinz would accept that empathy necessarily involves inhabiting another’s perspective. However, he does at least think that it sometimes will: he describes an “act of empathetic imagination, in which I experience, though vicarious emotions, how some action might have caused another person to feel.”  

Bloom and Prinz do both seem to include empathy in my sense without the scope of their criticism, then, although Prinz’s criticism may be also be intended to apply to phenomena distinct from the one that interests me.

Prinz’s stated aim is to show that “empathy is not all it is cracked up to be,” and that “Empathy as [he has] defined it may even be bad for morality.”  

Bloom is less circumspect about empathy’s moral badness: on his view, empathy is not only morally unnecessary, it is “morally corrosive.”  

For both of them, a critical first step toward putting empathy is its proper (much diminished) place is to show that the moral functions we might have thought to be the exclusive province of empathy may also be performed by other psychological mechanisms. As I mentioned above, Bloom and Prinz target empathy’s claim to an ineliminable role in the development and exercise of our moral epistemic competence, and also its claim to an ineliminable role in shaping and driving our moral motivation. Let us first review their reasons for thinking that when it comes to moral epistemic competence, empathy is dispensable.

Prinz begins by arguing that empathy is not itself constitutive of moral judgment. He points out that

---

we can empathize with others without approving of their attitudes as morally correct (in his example, a “recovering pedophile” can empathize with another pedophile’s lust while also judging it to be morally bad). Much of the previous chapter was devoted to an exploration of the distinction between what empathy guarantees (apprehension of the intelligibility of others’ emotions) and what it does not (judgment that those attitudes are correct), so it will come as no surprise that I entirely agree with Prinz on this point. Prinz and Bloom are also both concerned to show that empathy is not a causal precondition for moral judgment. They point out that experience tells us our judgments about the moral correctness of others’ attitudes can be formed without the aid of empathy. We can base our judgments on whether the other’s attitude conforms to a religious code of moral law, for instance, or on cold calculation about whether their attitudes and their accompanying intentions can be expected to maximize happiness. Bloom and Prinz underline that we also don’t need empathy in order to learn facts about others’ experience that are relevant to moral judgments. We can learn that others are suffering or offended through testimony, without needing to get inside their heads. I agree with Bloom and Prinz that in individual cases, both knowledge of others’ inner states and judgments about moral correctness can be secured without emotionally charged imaginative immersion in others’ perspectives.

After dispatching the possibility that empathy plays a critical role in moral judgment, Bloom and Prinz both target the suggestion that it is necessary for moral motivation. Here, too, the refutation proceeds primarily through counterexample. Prinz argues that we can be driven to act in “pro-social” ways through various non-empathetic means. For instance, we can be moved instead by guilt and the prospect of reward, such as “the pride that would come from our activism” fighting for the rights of minority groups (Prinz takes it that we can be reasonably confident that pride rather than

---

empathy is the prime mover when it comes to social justice activism, since empathy is “hard to muster for out-groups and collectives”). \(^{10}\) Other non-empathetic forms of moral motivation Prinz and Bloom point to include “repugnance,” as exemplified by the motivational orientation of the conservative opponents of gay marriage, and “cold logic and reasoning,” as in the case of an individual who donates a kidney because he judges that his life is not exponentially more valuable than that of a person needing a transplant. \(^{11}\)

Not all of these examples are convincing as evidence for the thesis that we can be morally motivated without empathy. Being motivated by pride is not necessarily the same thing as being morally motivated (if what one takes pride in is, for instance, the social status that one accrues as an activist). And repugnance at the thought of gay marriage hardly seems like a moral motivation. Perhaps that it how the conservative activists conceive of it, but there is a difference between believing oneself to be motivated by moral considerations, and actually being so motivated. That said, it does seem like empathy is not inevitably necessary for moral motivation. As has often been observed in the philosophical and psychological literature on empathy, high-functioning autistic individuals who are affected by serious deficits in their perspective-taking abilities nevertheless often demonstrate strong concern for justice and duty. Jeanette Kennett observes that such individuals seem to be natural Kantians. Perhaps the strong emphasis they place on ideals of “order and consistency,” as Kennett puts it, sets the moral motivational profile of high-functioning autistic people apart from the normal motivational orientation toward the good and right. \(^{12}\) Still, their existence gives the lie to the claim that empathy is required for any genuine moral motivation.

\(^{10}\) Prinz (2011a): 227.


\(^{12}\) Kennett (2003): 351.
In demonstrating that empathy is not required for the formation of moral judgments or for moral motivation, Prinz and Bloom take themselves to have shown that there is nothing morally significant that we can only secure through empathy. We can learn moral truths simply through testimony, and we can be morally motivated by, for example, Kantian reflection on the value of acting dutifully, so we don’t really need empathy at all in order to be morally virtuous. By itself, that intermediate conclusion licenses Prinz and Bloom to reject Shelley’s assertion that empathy is the great instrument of the moral good. They want to go further, though. Being empathetic, they hold, is not just one of many routes to virtue. Instead, it is a sure route to dangerous moral myopia.

Bloom objects that it is “a mistake, often a tragic one,” to think that being empathetic is part of being a good person, on three principal grounds. First, empathy is biased. We are more likely to empathize more fully with people who resemble us, who feel familiar, and who are attractive. Second, empathy is narrow. It inevitably focuses our attention on individuals rather than groups, and its scope simply cannot, as a matter of psychological fact, be extended to include even a small fraction of the teeming masses who populate the earth. Bloom regards these two features of empathy as problematic primarily because they result in a distorted distribution of the empathizer’s altruistic concern. He does not deny that empathy can prompt us to care, but he claims that it will be at the cost of making us care too much for some and not enough for others. Bloom’s third complaint about empathy is somewhat different. He claims that while non-empathetic rational compassion is a healthy feeling that seems to do no harm to the one who experiences it, empathy can be bad for the empathizer. Imaginatively immersing oneself in others’ painful situations can be

---

13 Bloom and Zaki (2016).
very stressful, and take a real toll on one’s psychological health. If empathy actually harms empathetic individuals, and if it is not necessary in order to secure any morally important goods, then it should not be promoted as part of a virtuous life.

The complaint about empathy’ narrowness is not obviously justified, because it does seem that we are able to empathize with abstractly characterized individuals, conceived of as representatives of groups. So, for instance, I might empathize with “a victim of Boko Haram’s kidnappings,” “a Californian almond grower,” and so on. If I can do that, then it is not as though empathy necessarily neglects the interests and concerns of groups. However, I will not object to the complaint that empathy is subject to bias. It is almost certainly true that, all things being equal, we empathize more completely, more often, and more easily with those whom we know better, or whose situations are more akin to our own. Empathy does sometimes have a tremendous reach. We can and do “bring home to ourselves” the anguish and joy of people separated from us by vast expanses of time and/or space, whose circumstances are quite alien to us (witness our ability to imaginatively immerse ourselves in stories about 14th century Veronese star-crossed lovers). Still, its most familiar and ready targets are our friends and neighbors.

I’ve reviewed the main points Prinz and Bloom make in their criticisms of empathy, and I’ve conceded that a number of these are highly plausible. However, even if we grant to Prinz and Bloom that empathy tends to be biased, coming to us more easily when the targets of empathy are more substantially like us, and even if we grant to Prinz and Bloom that (1) we do not in every instance need to empathize in order to form moral judgments, and that (2) we do not in every instance need to empathize in order to be morally motivated, we will not be obliged to accept their conclusion that empathy is morally otiose or even pernicious. One reason not to accept that conclusion is that Prinz
and Bloom overlook a moral good that we cannot secure through even the most refined of Kantian reasoning or the most reliable of moral testimony. I turn to that good now.

3. The non-instrumental moral value of humane understanding

Contemporary work on empathy’s moral significance has tended to focus on how empathy affects the one who empathizes. Rather ironically, it has had little to say about how empathy matters from the perspective of the one empathized with. Even more ironically, one of the few theorists to actually explicitly consider how the targets of empathy value being empathized with is Paul Bloom. Bloom asks: “Do people in distress want empathy?” This is a good question. Theirs is a perspective we need to pay attention to if we want to get the whole truth about how empathy and its humane understanding matters. Bloom answers that we might want others to empathize with us because this will increase the likelihood that those others will come to our aid.15 Or we might want our enemies to empathize with us so that they will suffer as we have suffered, thereby evening the scorecard. Perhaps those two motivations sometimes figure in our desire for others’ empathy.16 However, they are by no means the only reasons that receiving empathy matters to us. After all, we often desire the empathy of others who are neither our enemies nor in a position to aid us. In order to bring out an important reason that empathy matters to us as its recipients, one that doesn’t reduce to either a vindictive impulse or a desire for aid, let us look to an episode from E.M. Forster’s 1908 novel A Room with a View.

3.1. Empathy and humane understanding in A Room with a View

15 “Often we want others to feel our pain. After all, we know that feeling empathy for an individual makes you more likely to help them…So if I’m suffering and I want your help, I can try to evoke your empathy” (2016): 150.

16 I am not convinced that it is normal to want our enemies to suffer empathetically, though: when it comes to vindictive desires, don’t we more typically wish not that our enemies would imaginatively suffer alongside us, but rather that they would actually suffer the same losses they have inflicted upon us?
In the middle of the novel, the young protagonist Lucy Honeychurch finds herself in a remarkable position: she has been kissed by George Emerson, a man whose unconventional nature both attracts and disturbs her. This new experience comes at a moment when her whole worldview has been placed in flux by her experiences in the unfamiliar land of Italy. She is eager to bare her heart to her companion, her cousin Charlotte:

The luxury of self-exposure kept her almost happy through the long evening. She thought not so much of what had happened as of how she should describe it. All her sensations, her spasms of courage, her moments of unreasonable joy, her discontent, should be carefully laid before her cousin. And together in divine confidence they would disentangle and interpret them all.\(^\text{17}\)

However, Lucy finds that Charlotte is unwilling to pay attention to the real texture of her emotions. Charlotte comforts Lucy, but rather than trying to see how things are with her, rather than trying to appreciate what, exactly, struck Lucy as joyous, as disappointing, as calling for courage, she immediately sets about trying to plan how to cover up the whole affair. Lucy’s disappointment is acute. She describes Charlotte’s response as a “rebuff,” and as a “wrong not easily forgotten.”\(^\text{18}\) Her experience of injury in this case is juxtaposed with her “revival” near the conclusion of the novel, which occurs when Charlotte’s \textit{bête noir} Emerson Sr. tearfully gives voice to Lucy’s innermost feelings: “Now it is all dark. Now beauty and passion seem never to have existed. I know.”\(^\text{19}\) When he speaks thus, Lucy recognizes that Emerson’s is “the face of a saint who understood.”\(^\text{20}\) Being understood, it seems, effects a dramatic transformation in Lucy’s relation to herself and the world. She finds herself able to carry on despite a tormented heart and the knowledge that her choices, in particular her choice to break her engagement with the snobbish Londoner Cecil Vyse, will be

\(^{17}\) Forster (1922): 117

\(^{18}\) Forster (1922): 126.

\(^{19}\) Forster (1922): 310.

\(^{20}\) Forster (1922): 310.
received by many as shocking and lamentable.

Plausibly, Charlotte’s failures in her relation with Lucy go beyond a lack of empathy. One can, after all, at least listen to someone talk about their feelings without empathizing with them, and Charlotte declines to do even that. But it certainly looks like Emerson Sr. revives Lucy from a “deadened” position into which she has been cast largely by Charlotte’s “rebuff” by empathizing with Lucy’s predicament. Emerson Sr. is voicing a thought, or a perception, as if from Lucy’s point of view, and one has the sense that this expression is not issued from a position of intellectual abstraction. When he speaks, he seems to be looking out at the world through the lens of her despair and bewilderment.

Why does Emerson Sr.’s empathy effect Lucy’s revival? It is not that Lucy is revived because his empathy amounts to an endorsement of her emotion. When Emerson Sr. gives voice to her despair, he is not saying that Lucy’s emotion is correct. He is certainly not enough of a cynic to believe that there is no beauty or passion in the world. Nor is there any indication in the text that Lucy values the elder Emerson’s empathy because she expects that his empathy will incline him to provide further aid to her. Lucy is too ingenuous a character to go in for that kind of calculation. So the value of his empathy to her must lie elsewhere.

Here, rather surprisingly, I think Smith can help us out again. Even though Smith identifies sympathy with approving judgment, his most powerful descriptions of why we crave sympathy do not actually make reference to a desire to be approved of. He explains: “Nothing pleases us more than to observe in other men a fellow-feeling with all the emotions in our breast,” because when we “have found a person to communicate the cause of [our] sorrow…upon his sympathy [we] seem to
disburden ourselves of a part of [our] distress; he is not improperly said to share it with [us]” (TMS 15). The experience of having another join us in our feeling, Smith seems to suggest, is consoling because it relieves us from being alone with our grief, or disappointment, or joy. And the state of being alone in one’s own emotional outlook is one that is in itself at best unsatisfying, and at worst deeply painful: “the sufferer…passionately desires a more complete sympathy. To see the emotions of [others’] hearts, beat time to his own, in the violent and disagreeable passions, constitutes his sole consolation” (TMS 21-22).

To bring the pain of not being empathized with more clearly into view, let us contrast Emerson Sr.’s response to Lucy with one we could imagine another character having. Suppose, for instance, that Lucy were to say to her suitor Cecil Vyse “Everything seems dark to me now,” and suppose that Vyse were a person so irrepressibly cheerful in his thinking, and so lacking in imagination, that he couldn’t at all imaginatively work his way into Lucy’s grim perception. In his view, all is light and joy, and so her pessimism is beyond the bounds of what is humanely understandable for him. But suppose, also, that Vyse were to regard his own indomitable optimism as an epistemic handicap, much in the same way that a person with chronic pain insensitivity might regard their faulty nerves. Even though Vyse cannot see anything, his own situation or others’, as “all dark,” he might have it on good authority that “all dark” is a way that things can look to people, that it is in fact the appropriate way to see things sometimes. And he might regard Lucy as the ultimate authority on whether her situation has an inky cast or not. So, he approves of Lucy’s emotion, because he trusts her to get things right in this domain, even though he cannot empathize with it.

Could this response I’ve imagined for Vyse be substituted without remainder for Emerson’s engagement with her situation, and still be as effective a source of “revival?” Vyse has at least done
something more for Lucy than did Charlotte, who didn’t respect or even hear Lucy’s testimony about her feelings. Being trusted, as Vyse trusts Lucy, is a way of being respected as a valued source of information. But notice that there is still a kind of loneliness involved in some ways of being regarded as an authority. Imagine what it would be like if others regarded you as a kind of guru, but also couldn’t see where you were coming from. If everyone around you accorded you a great deal of epistemic authority, and took your word that the loss was devastating, the sculpture was beautiful, the goal was worthy, but nevertheless could not bring themselves to see things as you see them, it would be perfectly natural and reasonable to feel that you were in some troubling sense isolated from other people.

The act of grasping how another’s attitudes make sense is an intimate one. In order to succeed we must share the sensibilities of the other, at least to some extent. If my sense of what is outrageous is so different from yours that I cannot find my way into it, then no empathy will be possible. Understanding in the sense that concerns me means appreciating that the way the other looks at the world converges with one’s own way of seeing, even if it is not identical to it. This understanding is a powerful way of being close to another person, and closeness is something that social beings like ourselves crave, not for any further or deeper purpose, but just because it is valuable to us in and of itself. In short, I think, humane understanding is a way of being in fellowship. In Lucy’s case, it seems to me, Emerson Sr.’s expression of empathy effects a particularly dramatic transformation because the outlook he is empathizing with is, precisely, a lonely one. Lucy feels herself to be by herself in the dark, not tied to others by the connecting strings of passion, facing down the prospect
of ostracization from her community. But, in empathizing with her very alienation, Emerson Sr. offers her a relief from loneliness.21

Lucy Honeychurch’s story shows that being humanely understood can be very valuable to its recipients even when the one who humanely understands does not judge the recipient’s attitude to be correct. Sometimes, this may not be enough for the recipient of empathy. We might want others not only to imaginatively find their way into seeing the world as we do, but also to endorse that emotional perception of ours as correct. Let us call the object of desire in a case like this “understanding approval.” Understanding approval is a form of endorsement, as is the approval at issue in the case where one is treated as a guru. But there is a big difference between the uncomprehending endorsement of an attitude and an endorsement that comes on the heels of

21 The relief from loneliness I am describing here undoubtedly has something in common with Stephen Darwall’s concept of “being with.” Since Darwall is one of the very few contemporary philosophers to seriously reflect on empathy’s role as a constitutive feature of intimate relationships, it is worth pausing over how his analysis relates to mine. Darwall describes “being with” in this way: “Two people are with one another or together in the relevant sense, when they relate to each other or sense their mutual willingness to do so along with their mutual awareness of this mutual willingness. People who are thus together or with one another are open to one another and mutually aware of their mutual openness” Darwall (2011): 6. According to Darwall, empathy matters to mutual openness or accountability in part because “the projective empathy implicit in respecting the other takes them into the other’s perspective so that they can see themselves in the resenting way the other sees them” (17). That is to say, being with involves being willing to see oneself through the others’ eyes, such that one takes the other’s view of oneself seriously, and empathy is required for that operation. That is not quite the relevant dynamic in Lucy’s case: Emerson is not engaging with how Lucy sees him, but rather the world at large. But while “second-personal answerability” is the core dynamic with which Darwall is concerned, he also notes that empathy can matter to the quality of intimate relationships in other ways. Specifically, he claims: “We can adequately register them [another’s concerns] only by seeing his situation from his perspective and fully taking in the way things seem from his point of view. Only then do we really hear or “get” him.” As for what this “getting” amounts to, Darwall explains it will involve being “able to see by projective empathy that they [the concerns] are reasonable or at least understandable from his point of view. This too is a kind of affirmation we seek from friends and loved ones” (19). I am in complete agreement with Darwall that is very important for us that our friends and loved ones “get” us; in Chapter Three, I offered what may be understood as an account of what this “getting” consists in, and an explanation what exactly “getting” someone has to do with empathy.
humane understanding. It is only in the latter case that the empathizer appreciates the appropriateness of the attitude firsthand, for herself.

As it happens, understanding approval is likely to be a common outcome when efforts to empathize succeed. That is because (as I briefly remarked earlier) empathy and approval are often fellow travelers. Although there is a real and important gap between having an emotional response to some phenomenon and fully endorsing that response, we are, much of the time, powerfully inclined to treat as faithful to reality our own emotionally-colored presentations concerning how the world is. Full confidence in the accuracy of our own emotional representations is our default attitude. If this latter claim is surprising, it is only because we have a tendency, when thinking about emotions in the abstract, to focus on the ways in which they differ most sharply from mental states that we take to be more rational and objective. So, we call to mind burning rage that spreads itself indiscriminately, or we think of acute terror that cooler heads would dismiss as wildly overblown. But it is misleading to think of these as paradigmatic emotional experiences. Our ordinary experience of the world is valenced – this is boring, that enticing, this regrettable, that delightful, and so on. That is, ordinary experience is emotional, and we usually take those experiences at face value. We do not typically downgrade our trust in the representational deliverances of emotion until we have critically reflected upon the accuracy of the worldview the emotion presents. And this is just as well. A global policy of suspending trust in our own emotional experience could not be practicable. If we were to refuse to accept what is right before our eyes, emotionally speaking, we would have an awful lot of trouble making our way through the world.

Our own emotions are inevitably intelligible to us. Our acceptance of them as accurate evaluative perceptions is not inevitable, but that is the position we tend to default to. There is no reason to
think this tendency to default to acceptance will manifest only in the case of non-empathetic emotions. So, recipients of empathy can often expect their emotions not only to be humanely understood by empathizers, but also to be the subjects of understanding approval. Sometimes, we might crave both of those things. At other times, approval (understanding or otherwise) may be entirely beside the point: in a case where the recipient of empathy thinks her own emotion is not reflective of reality, it won’t make much sense for her to crave others’ approval of her attitude as correct.

Prinz allows that “We would all like to have empathetic friends,” that it is nice when our “near and dear” can empathize with us,” but claims: “it does not follow that empathy is desirable as a moral emotion,” because “an endorsement of empathy requires more than a warm fuzzy feeling. We need an argument for why empathy is valuable in the moral domain.” But, if I am right that we are pained by others’ lack of humane understanding, and if furthermore, I am right that be humanely understood, we need others to empathize with us, then facts about how we experience empathy are highly relevant to its moral status. Prinz should conclude the same, unless he would like to maintain that the relief of pain is a matter of no great moral consequence. In a large and important set of cases, perhaps the vast majority, it is morally good for us to empathize, not just because doing to will help us to apportion admiration and denunciation correctly (the benefit that would naturally fall out of Smith’s account of empathy as judgment) and not just as a means to promoting specifically altruistic instincts, either, but because empathizing is itself a way of responding to others’ needs.

Usually, when we discuss the moral significance of an epistemic good, the good is cast as instrumentally morally important, a means to performing well. Or, rather more rarely, it is

---

sometimes said that knowing *moral* truths is inherently morally valuable, that simply knowing the good is itself part and parcel of being good. The moral value of humane understanding that I’ve been discussing here is not instrumental, not is it just a matter of being appropriately aware of moral facts. Rather, it has everything to do with our relation to other people and their needs. The fact that humane understanding’s non-instrumental moral significance is relational means that it is sensitive to the needs of the individual target of empathy. Consequently, it is important to recognize the respects in which our desire to be humanely understood may be circumscribed or outweighed.

### 3.2. Variations in the need for humane understanding

Sensitive people can be intriguingly ginger about trying to empathize with others. People often say to the aggrieved: “I can’t even begin to imagine what that must be like.” That is less a report of a failed attempt to imagine, and more an acknowledgement that it would not be a good idea to try to imaginatively adopt the other’s perspective. It may be that we sometimes adopt this attitude just out of laziness, because it would take too much effort to do the imaginative work. But this is not always what is going on. The claim that we “can’t begin to imagine” can be an expression of a morally inflected hesitancy, a reflection of a sense that our efforts to empathize may be unwelcome. I will touch upon a few reasons why empathizers might not welcome efforts to empathize, or might welcome only efforts that are constrained in particular ways.

Some of us seem content with, or at least ambivalent about, the kind of isolation empathy relives. Hegel is said to have complained “Only one man has ever understood me. And he didn’t understand me.” Kierkegaard ups the ante: “People understand me so little that they do not even understand

---

23 Quoted in Heine (1892): 157.
my laments over their not understanding me.” A cynic, though, could suspect that Hegel and Kierkegaard, for all their lamentations, would not care to be “understood too quickly,” as André Gide puts it. Intellectual vanity may pull against the desire to have others easily access one’s perspective, and share in it imaginatively. A person who regards their own sensibilities as uniquely refined might not want a simple person to be able to humanely understand his outlook. This type of person might take a kind of pride in being set above and apart from others.

There are also reasons not to want to be humanely understood, or not to want to be humanely understood by certain persons, that are perhaps both more interesting and more respectable than mere snobbery. Anthropologists have observed that in certain cultural contexts, empathy is sometimes experienced as undesirably intrusive, something akin to mental trespassing. Jason Throop, a researcher who works on attitudes toward empathy in Micronesia, claims that “gaining empathetic insight into another’s subjective life” is not always what humans want from each other, and that our interest in being empathetically understood may be highly contingent upon our particular conception of the good life: “In Yap [a small collection of Micronesian islands], for instance, where the moral person is understood to be an individual who ideally embodies the virtues of self-governance, emotional quietude, and mental opacity, empathy and empathetic-like processes can carry with them more than their fair share of ambivalent feelings.”

---


26 Interestingly, this idea that empathy might be objectionably intrusive hasn’t tempted many philosophers (though see the reference to Nagel below).

Throop opposes empathy to mental opacity, which in this context is the trait of having one’s mental goings-on be well-concealed.28 Mental opacity seems to be particularly prized in the islands of the South Pacific, but even much more outgoing cultures make space for it to matter in certain contexts. We might also want others to respect our mental privacy for strategic reasons. As Thomas Nagel pointed out, too much in the way of revelation of inner lives could seriously undermine our willingness to even live near other people, much less relate to them as members of a moral community: “There is much more going on inside us all the time than we are willing to express, and civilization would be impossible if we could all read each others’ minds.”29

Throop’s description of the Yapese ideal of self-governance and self-sufficiency points to another reason why individuals might feel ambivalent about others’ empathy: although the conception of empathy I am working with is obviously distinct from that of pity, empathy can sometimes have a strong connection to pity, what we call “feeling sorry for.” When a person in a position of relative power or comfort becomes aware of another’s experience of being relatively diminished along some important dimension, the gulf between their relative standings becomes salient. A good-willed person attending to the difference between their standing and that of another person is thus liable to feel pity. Pity is a benevolent kind of looking down, but it is a kind of looking down nevertheless. Not all kinds of loving care are good for us, and this is one that people keen to preserve their pride might wish to avoid.

28 Note that this sense of “mental opacity” is distinct from the epistemic doctrine of mental opacity, which holds that others’ minds are unknowable.

The problems I have just mentioned have to do with the various reasons we might wish people *not to know* our own minds. Are they reasons to be particularly careful with empathy? Or just general reasons not to pry into other’s inner lives, ones that bear in just the same ways upon our other modes of knowing other minds? It does seem that when we are in the mood for mental privacy, we tend to be more put off by efforts to “really get inside our heads” than by, say, efforts to read our attitudes off of our gestures and behavior. But why should that be? I suspect that there are at least a couple of reasons why attempts at empathy may be experienced as more invasive than other kind of attempts to learn about one’s inner life.

First, there is the simple fact that we generally accept that our movements and expressions are prone to telegraphing our inner states, even when we would rather they not. We can’t really resent people for reading what is there to be read, whereas we might begrudge them their more effortful attempts to imaginatively adopt our points of view. Second, and perhaps more interestingly, empathy’s imaginative character does seem to make it particularly difficult to *contain*, such that invasiveness is minimized. Empathy has more of a tendency to “spread” than do other ways of knowing others’ minds. A contrast may help to make the point. Suppose that I want to find an answer to a particular question about your mental states: are you in pain? When I set out to learn about your thoughts by observing your behavior, and drawing inferences based on the usual connections between, say, grimaces and pain, it is relatively easy to restrict myself to seeking out answers to that question. Conversely, when one imaginatively entertains a worldview, picturing the other’s situation in detail, one is liable to make discoveries about the person’s inner life not strictly linked to the investigative questions one may have originally set for oneself. So, putting myself in your position, I apprehend the painful quality of your situation, but I might also find myself feeling a surprising bit of vicarious glee at the thought of that situation. I might then come to feel a bit embarrassed about that glee,
which might lead me to feel concerned that others are aware of my secret twinge of delight, and so on. This train of feeling might generate all sorts of ideas about your inner life, beyond the belief that you are in pain, including suspicions that you may be secretly enjoying displaying your pain, you are rather ashamed of that enjoyment, and so on. When we empathize, we set our emotional capacities to work, and our emotional sensitivities don’t always work in predictable ways. They have a habit of moving along unexpected associative paths, and turning up unexpected results. So, those with substantial imaginative abilities may find themselves uncovering surprising thoughts and attitudes, ones that those being empathized with might rather we weren’t privy to.

A last reason for ambivalence or concern about others’ attempts to empathize has less to do with the discomfort of being understood too well, and more with the fallout of failure to achieve humane understanding. When we “get” some outlook empathetically, we are bound to appreciate its intelligibility. We are also often inclined to regard it as correct. Conversely, when we attempt to empathize, and find we cannot, we are liable to be strongly inclined to think that there can’t be anything to what the testifier tells us. Claims that we can’t empathize often have a judgmental dimension. Think about what is conveyed when we say: “I will never understand why he would want that,” or “You say it’s worthwhile, but just can’t see it.” Or take this example from a Chicago Tribune Op-Ed: “So I will never get this sexting craze. I can't empathize with a teenage girl seeking acceptance by sharing photos of her breasts.” The literal claim being made is just that the author is not able to make sense of teenage girls’ sexting habits. But the heavy implication, here, is that there is no sense to be gotten, that the girl’s emotional outlook lacks intelligibility.

---

30 Albom (2014).
There is a non-accidental connection between our trying and failing to empathize with another’s attitude and our dismissing it as unintelligible and/or incorrect. A truly epistemically humble person might well think: whether or not I can imaginatively get myself to a version of the other’s attitude has no bearing on whether her attitude is correct, because I myself am (for instance) emotionally and/or imaginatively stunted. Still, this kind of humility is hard to maintain. We may be relatively content, sometimes, to skip the empathizing and let others put in the work when it comes to determining whether something looks (or is) beautiful, or just, or unsettling. However, once we do try to see at things from the other’s point of view, marshaling our emotional sensitivities and applying them to the task of figuring out whether the other’s evaluative perception is intelligible, it can be difficult to accept that there are aspects of the evaluative landscape that are beyond our ken. From the perspective of the one empathized with, others’ attempts to empathize may induce anxiety because of the risk that those others will fail in their efforts and that one’s outlook will therefore be dismissed as unintelligible.\footnote{I discuss this dynamic at greater length in Bailey (forthcoming), where I explore the relationship between empathy and trust in others’ testimony about evaluative properties.}

Given the potential complexity of our feelings about being empathized with, it is clear that empathizing in a virtuous manner will require sensitivity to the individual needs of the targets of empathy. In addition to the kind of concerns I’ve just described, the virtuous empathizer will need to be attuned to variations in the depth of understanding that will satisfy the other’s needs. One person might need for their specific emotion, with all of its detailed texture, to be understood by another: they need for the other to see how the loss of \textit{this particular} friendship calls out for sadness. For another person, such details may be irrelevant: their sense of isolation could be effectively relieved by a more coarse-grained empathy with an experience characterized more minimally as one of significant loss, and efforts to achieve a more detailed appreciation might be received as unhelpful.
or even intrusive. All that being said, the fact that the need for empathy is inconstant and context-dependent does not bar the conclusion that it is morally serious. Our need for love is similarly discriminating. We do not want or need to be loved by all people, in all ways, at all times, any more than we need to be humanely understood by all people, in all ways, at all times, but we do not think the need for love is for that reason un-serious or morally inconsequential, and we shouldn’t dismiss empathy’s value on such grounds, either.

3.3. Humane understanding and moral hazard

One might wonder whether there is another reason to be concerned about the morality of pursuing humane understanding, separate from recipients’ possible ambivalence about its value. I have frequently encountered this sort of worry about the proposal that humane understanding is non-instrumentally valuable: isn’t there something morally bad about empathizing with vicious perspectives, even in cases where we can reasonably expect that the one with the vicious perspective would very much like to be humanely understood? I don’t want to allow that there could be something morally good about empathizing with the vicious emotions of child molesters, serial killers, and the like, do I?

In response to this concern, the first point to stress is that putting in the effort to empathize with a vicious outlook will of course not always be the right thing to do, even if the bearer of that outlook is miserably isolated and craves humane understanding. Doing so might be incompatible with undertaking other more morally urgent courses of action. But that clarification does not, I think, take us to the heart of the worry. There are two ways of interpreting the core of the concern. The first is this: one might be worried that pursuing humane understanding of vicious outlooks could lead the empathizer astray, such that she herself becomes increasingly vicious. Efforts to grasp first-
hand the intelligibility of vicious outlooks such as, say, the vengefulness of the Dumas’ Count of Monte Christo, or the nihilistic misanthropy of Louis-Ferdinand Céline’s Ferdinand Bardamu, might encourage the incautious reader to see their own world through a vengeful or nihilistic lens, and even to endorse those ways of seeing as accurate. This worry is a legitimate one. There is the risk that in empathizing, one might lose track of the distinction between imagination and reality, or the distinction between how things appear and how they actually are, or both. So, people with a weak grip on the boundaries of imagination and/or the difference between what is correct and what is emotionally intelligible may be wise to refrain from empathetically engaging with objectionable viewpoints that they are particularly liable to adopt.

For at least some people, though, the concern about empathy with vicious outlooks is not so much that some of us may be susceptible to actually adopting and endorsing them as our own, but rather that empathizing with such outlooks will be morally sullying even if we retain a firm grip on the aforementioned boundaries. For basically good people, the risk that we will imaginatively find our way into wildly cruel outlooks is not great: the idea that skinning kittens is enjoyable is not liable to make sense to us. But if we could find our way, should we? We might frame the question in terms of ideals: will a greatly virtuous person be the kind of individual who would say, with Terence, “I am a man, I consider nothing that is human alien to me”? Will they be able to appreciate what is intelligible in others’ murderous ambitions, callous cruelty, and selfish scheming? Or will the greatly virtuous person be more like Dostoyevsky’s Prince Myshkin, a “holy fool” so pure of heart that others’ dark desires baffle him entirely?

---

32 Bardamu is the murderous, spiteful, shockingly uncaring anti-hero of Céline’s very upsetting novel *Journey to the End of the Night.*

33 Nietzsche’s lines about fighting monsters and gazing into the abyss have lapsed into cliché, but they still do capture a real possibility.
Here is a point in favor of the former possibility. There is something unappealingly prudish (or, at best, frustratingly useless) about a person who is, as it were, “too good” to empathize with others’ evaluative outlooks, when those outlooks are morally bad in a venial sense, and when the bearer of that outlook desperately needs to be humanely understood. Imagine: you are envious of your colleague’s many accomplishments. You know that you oughtn’t to feel this way, and you berate yourself for it, but some part of you (most of you, if you’re being honest with yourself) wishes that she might fail at something, for once in her life. You are ashamed of your emotion, and you are very hesitant to tell anyone how you feel, but you still find yourself miserably dwelling on your colleague’s superiority. Imagine you confess your envy to me. You ask if I can see where you are coming from. I say that I am sorry, but I must decline to try and empathize with you: I should not pollute my mind with unlovely envious thoughts. I take it that you would be within your rights to resent me for my unwillingness to see things from your point of view. There is something maddeningly priggish about my avowed priorities. If, alternatively, my problem is not an unwillingness but an inability (envy is just not part of my nature), resentment will not be warranted but disappointment might be. You needed someone to get into the emotional trenches with you, and help you find your way out. Instead, you found yet another person who is naturally better than you, and who for that reason is not in a position to empathetically share your feelings, and to relieve you from the burden of worrying that your emotions are so monstrous as to be unintelligible to ordinary human beings.

I will not make the case, here, that the most virtuously empathetic person is one who is willing and able to imaginatively take on the perspectives of unusually sadistic or otherwise awful people. I am not confident that this is even the right conclusion to reach. If there is something that seems right in the story I have just told about the unwillingness or inability to empathize with mildly bad (but
eminently human) outlooks, though, then we may have some moral reason, at least some of the time, to dip our toes into morally suspect evaluative outlooks. To refuse to do so could sometimes amount to an undue prioritization of one’s own purity of mind over the real needs of others.

3.4. Humane understanding and friendship

One attraction of the view that being humanely understood is desirable for us because it relieves a painful form of isolation is that it can explain how empathy can matter at a distance, while also fitting with our sense that empathy is especially important in the context of close relationships. The thought that someone understands us, even if they are a stranger to us, can be a consoling one. But when it comes to the people who we share our lives with, the pain of being regarded as alien, of not being understood, is that much greater. Humane understanding is an important part of our most intimate relationships. In fact, recognizing the role of humane understanding in friendship may actually help to resolve a longstanding philosophical disagreement about the nature of friendship, in a way that should appeal to those on both sides of the debate.34

Aristotle suggests that friendship involves finding the same things pleasant and painful as one’s friend does. He claims that individuals unequal in virtue could not be true “character friends,” for “how could they be friends when they neither approve of the same things, nor find the same things enjoyable or painful? Nor even with regard to each other will this be so, and without that they could not be friends” (EN 1165b14-35).35 Some philosophers second Aristotle’s opinion on this point. True friends must share values. Richard White claims that “Aristotle is correct, friendship implies solidarity, or a basic sharing of values and perspectives and a similar sense of what is important and

34 Thanks to Jennifer Whiting for discussion on this point.

35 References to the *Nicomachean Ethics* are to Crisp’s (2000) translation.
what is worth doing.” Elizabeth Telfer similarly remarks that rather than mere liking, friendship requires a bond of similarity: “The bond may be shared interests or enthusiasms or views, but it may also be a similar style of mind or way of thinking.” Other philosophers sharply criticize the claim that friendship requires the sharing of values. Dean Cocking and Jeanette Kennett argue that the “mirror view” of friendship, according to which true friendship is distinguished by “the great extent to which we see ourselves in the close friend,” should be rejected. They claim that it is obviously false because as a matter of fact “people who are very dissimilar can be very good friends.” Of Mice and Men’s Lenny and George are very dissimilar in their character and values, Cocking and Kennett claim, but that does not prevent them from being great friends. In fact, they go on to say, such dissimilarities can be a ground of friendship. In the case of a friendship between a cautious friend and a reckless one, “the reckless one knows that she could never resemble her friend in recklessness, yet she is attracted by this aspect of his character. The reckless one, while remaining quite reckless, regards with affection his friend’s caution.”

So, who is right? Is friendship not only compatible with but also sometimes sustained by differences in evaluative outlook? Or do such differences always detract from the perfection of a friendship, other things being equal? Both proposals have some appeal. Happily, recognizing the role for humane understanding in friendship will help us to find a middle ground between them.

---


37 Telfer (1970): 227. Telfer actually immediately goes on to say that this bond “makes for a high degree of empathy.” It is unclear, though, what sense of empathy she has in mind.

38 Cocking and Kennett (1998): 507. Other examples Cocking and Kennett cite include Jack Spratt and his fat-eating wife, plus Darcy and Bingley of Pride and Prejudice.

To Cocking and Kennett it should be readily conceded that friends don’t have to be emotional carbon copies of ourselves, bearers of cares, desires, and feelings identical to ours. In fact, a carbon-copy “friendship” could only be strange and pathological. Just think: If I actually loved your family and partner just as you do, I would be less a friend, and more someone who had confused myself for you. And some minor forms of mutual inscrutability can indeed be tolerable in a relationship, something to laugh over and tease about: I find your passion for comic opera baffling, and you feel the same way about my love of Balkan brass music. However, what friendship does require is that each person can appreciate (with a little imaginative effort) where the other is coming from, when it comes to those aspects of our emotional experience that are closest to the core of who we are and what we care for. That is to say, friendship requires empathy with the other’s outlook.

I do not need to actually love your partner as you do in order to be your friend. If anything, my loving your partner in this way could be a barrier to our friendship. Nor do I need to actually value your career or your other major life-orienting activities as you do. However, the following thought rings true: if, despite my best imaginative efforts, I can only apprehend your beloved partner as an infuriating bore, or if you cannot appreciate how the career I am devoted to could look like anything other than a dreadful, wasteful, slog, then our friendship will most likely come under real strain. I think there are two reasons for this. First, our friends are supposed to relieve us from the pain of being alone, in multiple senses. A good friend spends time with you, participates in activities with you, listens to you. But even someone who sits beside us and listens to our problems can remain at quite a distance from us, psychologically speaking. In fact, we are liable to feel especially alienated from a person who is attentive and physically present but still unable or unwilling to interpret our concerns and yearnings as intelligible, since it will be harder to blame this failure or unwillingness on other, intervening factors. A second, auxiliary consideration is that we
look to our friends to help us refine our sense of what our emotions are about and think through the ways in which they might need to be adjusted to reflect the actual evaluative facts on the ground. A companion who is unable to achieve humane understanding of one’s emotion is not likely to be able to be very helpful in these respects, since she will not appreciate what exactly *feels* right about the emotion that is to be refined or redirected. So, someone who is consistently unable or unwilling to humanely understand our emotional outlook will, in the end, not be much of a friend, although they may continue to be someone we care about a great deal.

The role of empathy in friendship is a large part of what makes friendship different from a relation that is simply one of mutual respect or admiration. We can respect or admire others without much in the way of humane understanding of their attitudes. In fact, the absence of such understanding can be part of the awe that distinguishes our admiration for gurus, sages, and heroes. The forbearance and fortitude of such figures strikes us as superhuman, worthy of our highest esteem but so great that it is beyond the scope of our empathetic comprehension. Friendship is about more than approval, though. Or, rather, it may not be very much about approval at all; the fact that you see the world in the same very gloomy light that I do might help to bind us together in friendship even if neither of us approves of our Eeyore-ish outlook as an accurate reflection of the world’s evaluative features. If I can’t bring myself to see the world as you do, even in imagination, then I can admire you, like you, and enjoy your company, but there will still be a psychological distance between us that is inimical to friendship’s characteristic intimacy.

---

40 See page 154ff for a discussion of the progressive articulation of emotions.

41 To revisit the case of Lenny and George: the ranch-hands were undeniably close in some important respects. They lived together, and they very much cared for each other. Still, it seems fair to say (contra Cocking and Kennett) that their relationship was not fully a friendship, but something more like the relationship between a parent and child. Recall, Lenny’s developmental handicap meant that he did not understand much of what George said to him, and he was persistently bewildered by George’s frustration and concern.
Earlier, I granted that empathy might be biased or otherwise capricious. Indeed, when it comes to mediocre moral agents (which is to say, most of us), it is practically inevitable that not everyone will receive the empathy from us that they need or deserve. But if empathy is in fact the only way to relieve a particular sort of pain, then something would undoubtedly be lost if we were to try to curb or squelch our empathetic efforts. I have not tried to prove here that the heft of that potential loss definitely outweighs concerns about our empathetic efforts’ susceptibility to bias, but I do want to underline that something serious would be missing if we gave up on empathy. If we were to follow Prinz and Bloom’s counsel to suppress empathy in favor of cold reason, we would end up giving up on a central component of friendship, as well as an important source of comfort to many suffering people. Concluding that we should still try to do away with empathy because of concerns about bias feels a bit like insisting that we should try not to love anyone on the grounds that we cannot love everyone and that our distribution of love is subject to bias.

4. Empathy and Moral Epistemology

So far, I have been concentrating on making the case that empathy is good for a very different kind of reason than the ones Prinz and Bloom consider. I now want to return to the kinds of potential benefits of empathy that Prinz and Bloom do explore, and show that empathy makes a more real and serious contribution in these domains than the critics of empathy acknowledge. First, let us turn to empathy’s significance for moral epistemology, its role in securing the knowledge or understanding of the right, the good, and other moral properties.

As a reminder: Prinz and Bloom rightly make the anti-Smithian point that judgment of attitudes as correct or incorrect can come apart from empathy. It is not as though empathizing with another’s
perception of something as offensive or cruel or what have you actually requires us to judge it to be so, and in individual cases we can form judgments about the propriety of others’ emotions without needing to empathize. However, we can grant those points and still maintain that empathy is in the long run essential to the development of our moral commitments and sensibilities. Over the course of a person’s life as a continually maturing social agent, empathy functions as a critical means by which we are able to extend and refine (1) our reflective judgments about evaluative properties and (2) our emotional sensitivity to these properties.

Let us first consider empathy’s impact upon the evolution of our moral values and commitments. As agents who aspire to virtue, we need to attend to the presence or absence of evaluative properties that are, generally speaking, complex and finicky. Whether something is beautiful, or funny, or nasty or contemptible can hang on even the smallest shift in circumstance or detail. Give it too much of a cutting edge, and a joke can shift from being amusing to merely caustic. Take it upon yourself to casually underline the goodness of your deed, and it can turn from a generous act to an irksome one. Moral maturation is in large part a process of discovering that the borders between the helpful and the hurtful, the virtuous and the vicious, cannot be correctly drawn with the aid of a few simple maxims or rules of thumb.

Often, this maturation is driven by emotional experience. Our emotional sensitivities can outpace our reflective beliefs about the boundaries between evaluative properties.\footnote{A number of philosophers have convincingly argued for this claim. See e.g. Jaggar (1989) and Helm (2000).} So, for instance, we might initially think that the kindest thing to do, when another person is grieving, is always to try to cheer them up, on the grounds that feeling sad is an awful experience and it is kind to prevent others from having awful experiences, if we can. However, upon experiencing a serious loss, we might
discover that others’ efforts to make us happy feel more hurtful than helpful. We might find that what we need is not to be cheered, and we might experience others’ efforts to change our feelings as slighting rather than kind. If we are reasonably good moral learners, this discovery about how things can look from the perspective of a suffering person will lead us to revise our thinking about the border between a kind and helpful encouragement toward optimism, on the one hand, and a wounding failure to take another’s sadness seriously, on the other.43

In actually finding or placing ourselves in new situations, we are able to make morally relevant emotional discoveries of this sort. When we empathize, we imaginatively place ourselves in new situations. Does imaginatively engaging with possible scenarios afford the same possibilities for this kind of discovery that actually being in a new situation does? People often say of many different kinds of experiences that one cannot understand what it like to [lose a partner, forgive an enemy, go into exile, eat a pineapple pizza] unless one has actually “lived through it.” On behalf of the thought that imagination is no substitute for experience, this should be said: our pre-conceptions about the relevant values, and our predictions about what we are likely to feel, may color our imaginative recreation of, and emotional response to, situations that we have not actually found ourselves in. For example, in the imaginative analogue of the case I just described, my standing belief that it is kind to cheer sad people up, and my related expectation that I would probably appreciate being cheered up, may lead me to build into the imagined situation features that seem to especially invite such an appreciate response. These beliefs and expectations may also squelch any resentful emotional response to the imagined situation: If I am convinced that cheering up is the kind thing, I may be (unconsciously) inclined to suppress any welling up of resentment in response to the imagined situation.

43 Of course, we might also decide that our emotional response is distorted, not reflective of the actual presence or absence of evaluative properties, and so not a reason to shift our evaluative commitments. But if we accept that our emotional sensitivities can sometimes be more accurate than our reflective beliefs, it will not always make sense to reject emotional presentations in this way.
situation. Of course, expectations can have a distorting influence upon our emotional responses to the situations we actually find ourselves in, as well. My strong anticipation that I will find something enjoyable can, for instance, have an influence upon how enjoyable I actually find the experience to be. But because we exercise a creative control over our imagination that has no real analogue in the case of lived experience, it makes sense to think that our assumptions and expectations will exert a particularly strong influence over our emotional engagement with imagined scenarios.

That said, our emotional responses to vividly imagined circumstances can sometimes take us by surprise, and lead us to revise our considered views about moral evaluative properties. A classic parental tactic illustrates the point. A child who plays roughly with a smaller, shyer child might initially think that there is nothing wrong with her behavior: rough playing is fun and good as long as no one means to hurt anyone else, and no one says that they don’t want to play. She might even predict that she wouldn’t mind it if someone tried to play that way with her. But, when her parent helps her through an exercise of empathy (“How would you feel if you were small, and someone who was much bigger, louder, and stronger grabbed you? Would that be fun, or scary?”), the rambunctious child might nevertheless find that she experiences a prickle of anxiety, and grasps how her roughhousing could feel threatening. I suspect it is no accident that this (anecdotally) effective exercise is a guided one. The parent helps to determine what, exactly, is to be imagined, and highlights possible alternative emotional responses. The constraints the parent imposes on the child’s perspective-taking efforts reduce the availability of imaginative responses that would more in line with the child’s initial evaluative judgments (such as: “I wouldn’t mind, because we would just be having fun”).
One’s ability to make emotional discoveries about moral properties via empathy will likely largely depend upon how disciplined one’s empathetic efforts are. In the case just discussed, the parent helped the child to set her previous beliefs about the propriety of roughhousing well off to the side, and focus on responding emotionally to the imagined case. More mature moral learners can serve as their own parents, as it were. We can try to ensure that when we ask ourselves “how would things feel, from that position?” we are not just making a prediction based on what we hope or expect to be true, or suppressing emotional responses that don’t fit with our previous ideas about the helpful, the hurtful, the kind and the cruel. It is not always easy to be genuinely open to the possibility of emotional discoveries about moral properties, either in or outside of the empathetic context. Still, the moral educational benefits of this kind of openness are real. Even a person who leads an unusually varied and adventurous life can only actually encounter so many circumstances first hand. We can only literally live through a tiny sliver of possible human experience. Our ability to empathize multiplies many times over the circumstances with which we can emotionally engage. It therefore vastly expands our opportunities for uncovering problematic crudity in our beliefs about when and why things are sad, awful, honorable, admirable, and so forth.

I have been considering how through empathy, we might deploy our emotional sensitivity to evaluative properties in a way that motivates revision of our standing moral beliefs and commitments. Empathy also allows us to further develop our emotional sensitivity to evaluative properties. This latter proposal has actually already come in for some criticism. In a very engaging exploration of empathy’s significance for moral epistemology, Justin D’Arms suggests that while emotional contagion (which he categorizes as one form of empathy) can serve as a very effective means for the evolution of our sensibilities, the same is not true of “simulation” (an activity that corresponds to empathy as I have defined it). He explains:
In a simulation, the simulator brings her own sensibilities to bear on another person's situation. If this produces new feelings in her, it is her own sensibility that explains the presence of these feelings. The model's position [which is to say, the position of the target of empathy] only determines the nature of the inputs on which the simulator exercises her sensibilities. Contagion does something more radical. It gives the observer vicarious access to the sensibilities of another.44

D’Arms proposes that I can access the sensibility of another through emotional contagion in the following way. At first, I might hear a comic’s joke without feeling at all tickled by it. If my friend is laughing, though, I might I automatically mimic my friend’s expression of amusement. Then, relying upon environmental cues, I might cast about for the source of my friend’s amusement, an amusement which I have now been “infected” with. And once I identify that source (the joke), I might end up taking that joke as the target of my contagiously-derived amusement. D’Arms claims that this is quite a normal way of coming to shift one’s evaluative outlook: “it is a familiar phenomenon that people who wish to take a stern attitude, and who might be disposed to do so on their own, can be drawn in by a friend's amusement and find themselves laughing at something they would disapprove of under other circumstances.” 45 By contrast, when we can’t “borrow” others’ sensibilities through the mechanism of contagion, our ability to see things in a new emotional light is effectively non-existent. If we are just asked to imaginatively put ourselves in the place of an amused person, and determine whether we can indeed empathize with that person’s amusement at some joke, the only emotional sensibility we can bring to that exercise is the one we already have.

I am less optimistic about the radical transformative power of emotional contagion than D’Arms is. It seems to me that in determining the intentional object of our contagiously acquired amusement (or other attitude) we will tend to cast about for objects that are consistent with our pre-existing

sensibility. So, for instance, if everyone is amused by the racist stereotype at the heart of a joke, but I do not appreciate racist humor, I might instead take the object of my contagiously-acquired amusement to be, say, the funny voice in which the joke is told. But I do not want to linger over the question of what kind of epistemic contribution contagion can make, since my concern is with the impact of imaginative, emotionally charged perspective-taking.

D’Arms is right about the limits of perspective-taking’s transformative power to this extent: if we simply ask ourselves how some circumstances would look from the perspective of some individual (how some compliment would look to our co-worker, for instance), and seek to answer that question solely by imaginatively placing ourselves in her position, our efforts will just involve exploiting our own existing sensibilities. We will be relying on our pre-existing sense of the pleasing, the offensive, and so forth to answer this question, and, consequently, there does not seem to be much room in a scenario like this one for the evolution of our sensibilities. However, D’Arms’ analysis passes over another way distinct from emotional contagion in which others’ actual attitudes and opinions can inform our empathetic efforts. When empathetic effort is combined with trust in others’ testimony about evaluative properties, our perspective-taking efforts can indeed lead to the evolution of our sensibilities.

To see what I mean, lets consider the case of someone who says to me that her supervisor’s compliments are demeaning– his remarks to her are subtly barbed, and communicate a deep-seated disdain. I might initially respond to this testimony with puzzlement. It’s usually nice to receive a compliment, after all, and the remarks of the boss don’t fit the models of derisive communication with which I am familiar. But if I trust my acquaintance as an honest testifier who is likely to be a good detector of evaluative properties of this sort, then I will have a solid epistemic reason to look
again.\textsuperscript{46} That is, I will have reason to test my imaginative and emotional capacities, to try and stretch them enough to see for myself something hurtful in the supervisor’s communication, such that my emotional perception is, at is were, keyed to the evaluative features I have (defeasible) testimonially grounded reason to believe the situation has.

We cannot reasonably expect that this kind of imaginative exercise will regularly effect a very dramatic transformation in our own outlook, precisely because the extension that can be achieved in this way must work with the materials of our own existing sensibilities. Even very rich imaginative immersion in another’s situation is unlikely to allow a thoroughgoing optimist like our Cecil Vyse to suddenly acquire a keen attunement to the gloomy and desolate, for instance. Still, when it is guided by our trust in testimony, our capacity for humane understanding can be deployed for the refinement and extension (if not the wholesale re-invention) of our emotional attunement to the kind, the outrageous, the beautiful, and the demeaning. This extension can often be helped along by the use of analogical thinking. So, for instance, if I trust my acquaintance that that her bosses’ compliments are demeaning, but I do not immediately see how they could be offensive, I might look closely for ways in which his compliments resemble comments or gestures that have bothered me, despite their superficially benign appearance. Recognizing these points of resemblance can help to transform our emotional apprehension of the other’s situation.\textsuperscript{47}

In this section, I have briefly outlined two respects in which our ability to empathize might contribute to our knowledge and understanding of, as well as sensitivity to, morally relevant

\textsuperscript{46} I have in mind the empathetic analogue of the kind of looking again Iris Murdoch so memorably captures with the case of M and D. See Murdoch (1970): 18-19.

\textsuperscript{47} The extension of emotional sensitivity to evaluative properties through analogical thinking is a major theme in the work of Mengzi. See McRae (2011) for a helpful discussion.
evaluative properties. These two respects are complementary. On the one hand, empathy can help us to correct our beliefs about moral properties by exploiting the fact that our emotional sensitivities are sometimes more subtle and precise than our corresponding reflective beliefs about evaluative properties. Our emotional empathetic responses to imagined scenarios can take us by surprise. In cases where we are willing to trust those emotional responses, they can lead us to revise our judgments about the good, the bad, and the ugly. On the other hand, there are cases where our emotional sensitivities need to catch up with our evaluative beliefs, instead. We might initially find that another person’s appraisal of their situation does not resonate with us. But if we believe their testimony about how the situation is, if we trust them the remark is offensive, or the joke is funny, or what have you, then we can try to look again, to see if through sufficiently careful attention and detailed, vivid imaginative effort we can work our way into an emotional apprehension of the other’s situation that matches their own. If we succeed, we will then not only know, second-hand, that things are the way the other says they are—rather, we will secure a first-hand appreciation of the remark’s offensiveness, the joke’s humor, and so on.

Whether the moral epistemic gains secured through a particular instance of empathizing are more on the side of evaluative belief, or instead more on the side of emotional sensitivity, it is true that discipline and guidance have a real impact on how significant those gains are. Absent the guidance provided by others’ testimony, or the discipline required for genuine openness to the possibility of new and surprising emotional experience, our imaginative and emotional efforts are liable to tend toward confirmation of the status quo. We will lapse into relying upon our old beliefs about evaluative properties, or we will make little effort to stretch our capacity for emotional perception. That said, even someone who is not an especially talented or careful empathizer will have a real epistemic advantage over the individual who cannot or will not empathize at all. The moral learning
of the latter individual will be hampered because the realm of experience from which she can learn through first-personal emotional engagement will be very narrowly circumscribed.

5. Empathy and moral motivation

At long last, it is now time to return to the dimension of empathy’s moral significance that has most interested the majority of philosophers and psychologists working on empathy. Recall, the view of Hume and his philosophical inheritors was that empathy matters morally because it dramatically changes the way we care about other people. I have accepted that claim, in a sense: I have argued that the pursuit and provision of humane understanding is itself a way of caring for other people. However, I have rejected the Humean notion that empathy enhances care by extending our primitive concern for our own weal and woe to encompass the pains, pleasures, and fortunes of others. I have also granted to empathy’s detractors that it is not necessary for moral motivation, in both of the following respects: first, it seems that in individual cases, we can be moved to act morally without empathizing, and second, a robust Kantian concern for justice and duty seems to be sustainable even in cases where agents’ empathetic abilities are unusually limited. Finally, I have also pushed back on the tendency to equate empathy’s moral significance with its contribution to moral motivation by highlighting the fact that often, others’ desires to do right by us aren’t enough: we need others’ moral concern, but we also need their humane understanding. What is more, in addition to itself constituting a morally important way of caring for others, empathy helps us to refine and revise our moral commitments and to develop our emotional sensitivity to evaluative properties.

With all that being said, though, empathy can also have a moral motivational impact in that empathy makes it very difficult for us to treat people in particularly immoral, callous ways. This is not, I think,
empathy’s most significant or interesting contribution to our moral lives, but it does merit consideration. The ways of describing empathy’s motivational significance that I canvassed in Chapter Two do not include a special role for empathy’s characteristic epistemic function. Recall, Hume’s account, and the accounts that take after it, portray empathy’s epistemic and motivational functions as largely independent. Recently, there have been a couple of attempts to explain empathy’s contribution to moral motivation, and more specifically to altruism, that do give more of a pride of place to what their authors identify as empathy’s contributions to knowledge or understanding. Both Karsten Stueber and Jonathan Deigh have independently claimed that empathy with an individual entails an appreciation of that individual’s nature and/or attitudes that is incompatible with certain morally bad motivational stances. I will argue that there is something promising about the form of these arguments. In order to get at the way that empathy shapes our moral motivation, it will be helpful to think about the kind of recognition of others that empathy necessarily involves. However, both arguments conclude with claims on empathy’s behalf that are too grand to be defensible. Contra Stueber, it is not the case that in empathizing, I must conceive of the other as “having as much value as my own self,” and contra Deigh, it is not the case that empathy necessarily involves seeing others’ purposes as worthwhile, no matter how “mature” that empathy is.48

When it comes to the recognition of morally salient features of others and their inner lives, features whose acknowledgement imposes limits on the ways in which we can be moved to treat others, empathy makes a rather more modest but still important contribution. We cannot both empathize with someone and see them as unfeeling, as unthinking, or as lacking a normative perspective on the world. Consequently, empathizing with someone is at least not compatible with treating them as a

mere object. Typically, it will also not be compatible with treating the target of empathy as a mere passive observer of the world, rather than an agent capable of deliberation.

Stueber’s suggestion about what the empathizer’s moral stance toward others must be is explicitly presented as “speculative.”49 It is much less detailed than Deigh’s, so I will address it first. According to Stueber, “the basic cognitive stance of our folk psychological practices” is one “where we understand each other as rational agents acting for reasons.” This fact has consequences for our treatment of others, because “[i]n trying to re-enact another person’s reasons I fundamentally accept the other self to be like me and as having as much value as my own self.”50 The declared focus of Stueber’s analysis is our apprehension of others’ rationality through empathy, whereas I have suggested that empathy is most directly concerned with the intelligibility of others’ outlooks. However, it is fair to say that Stueber has in mind the same basic phenomenon, an imaginative recreation of the others’ perspective that is emotionally engaged, and that centrally involves getting a grip on the normative status of others’ evaluative outlooks. Stueber’s first claim here about what imaginative re-creation entails, namely that the effort to re-enact another’s reasons requires acceptance that the other self is like me, is a sound one provided that “like me” is understood in the right way. As I’ve argued in 3.1 of this chapter, in order to succeed in our pursuit of humane understanding, we must share the sensibilities of the other, at least to some extent. It would make no sense at all to try to empathize with someone if I were convinced that their emotional apprehension of the world could not at all coincide with mine. The greater the overlap in our sensibilities, the more robust and complete my empathy with them is liable to be.

Stueber does not elaborate upon the connection between empathetic recognition of another as “like” me and the attribution of equal value to the other, nor does he explain exactly what sense of “equal value” he has in mind. Suppose we understand “having just as much value” to mean something like “being equally as deserving of respect, care, and/or admiration” (which seems like a plausible enough interpretation). In that case, the idea that the other is just as valuable as me does not seem to be necessary for us to undertake or succeed in empathizing. We can grant that one possible reason for believing oneself to be more valuable than others, the thought that one’s sensibilities are uniquely refined, will put up barriers to empathy. Characters who believe themselves to be uniquely sensitive to evaluative features of the world, the kind of snobs I briefly discussed in section 3.2 above, might decline to even try to empathize, on the grounds that others’ sensibilities are so crude as to be empathetically inaccessible to refined creatures like themselves. But I might also believe that others are less valuable than I am for reasons that have nothing to do with my degree of emotional sensitivity to the world’s evaluative features. Perhaps I simply believe that I am uniquely precious, or even divine. I may therefore consider you to be very much like me in terms of our appraisal of the properties in the world, and thus a suitable target of my empathy, without for all that thinking that you are as valuable as I am. Even quite a robust empathy with another person may be compatible with the enduring belief that we are not equally valuable in some basic, critically important respects. I could even empathize with the bemusement or frustration that my conviction that I am God’s special gift to humankind inspires in you, without thereby compromising the strength of that conviction (from my perspective, this would be a case of empathizing with a mistaken attitude). In short, we cannot both empathize with another and think that we have nothing in common with that other, but we can both empathize with another and believe that they are not equal to us in value.
Let me now turn to Deigh’s more fully articulated attempt to explain why empathy is incompatible with what he labels a “psychopathic” moral orientation, characterized by “insusceptibility to moral feeling and moral motivation.” He begins from the thought that a person could have the following perfectly intelligible, non-contradictory attitude: Other people are both rational agents and instruments of my will, and while they care about their ends, it is only my ends that truly matter. For Deigh, the mere knowledge of others’ rational agency or personhood does not by itself generate a reason to have any non-instrumental regard for them. However, Deigh contends that the introduction of perspective-taking imagination will upset this coherent, if very alien, picture of value. Deigh explains:

Of course, if she [the psychopath] were to imagine having the purposes that another person has, then she would see how purposes other than her own can matter too. That is, if she regarded these purposes as this other person would, she would then see that they were worth pursuing and, having no reason to count them as less worthwhile than her own, would conclude that she and others ought no more to interfere with his freedom and well-being than he and they ought to interfere with hers.52

Deigh characterizes the kind of empathy that necessarily elicits this conclusion as “mature empathy.” It is a form of empathy that we develop over time. The empathy that children start out with is limited in scope, being concerned only with individual feelings or sensations. But as a child develops more sophisticated imaginative abilities, “it becomes capable of a maturer empathy in which the recognition of others as architects and builders, so to speak, of human lives is more pronounced. In taking another's perspective, it sees the purposes that give extension and structure to the other's life and sees those purposes as worthwhile, as purposes that matter. In this way it comes to recognize others as autonomous agents.”53

53 Deigh (1998): 760. In case it is not totally clear, the “it” in this passage from Deigh is a child.
On Deigh’s account, the moral development achieved through empathy has two fronts or faces. First, there is our maturing appreciation of the normative features of others’ purposes: through empathy, we are supposed to increasingly recognize others’ purposes as mattering, and worth pursuing. And second, there is our maturing appreciation of the normative features of others’ natures: through empathy, we are supposed to develop a sense of others as the autonomous architects of their own lives. I have reservations about the ambitiousness of Deigh’s claims on both fronts, although I do accept the broader insight that empathy requires us to see other creatures in a certain light, one that will influence how we can be moved to treat others. Let me first address the thought that empathy involves apprehending that the other’s purposes are worthwhile.

Empathy undoubtedly involves appreciating how the other’s purposes look worthwhile, from the other’s point of view, provided of course that the other does see their own purpose as worthwhile. If, to return to an earlier case, I cannot see anything intelligible in your plan to get married to the Eiffel tower, because I cannot see anything in the tower as inviting romantic devotion, then I will not have succeeded in empathizing with you. But there is a difference between appreciating firsthand how your purpose looks worthwhile and endorsing your project as worthwhile, full stop. Endorsing your plan is no part of empathy. I might deeply empathize with the emotions that inspire and inform it, and yet still judge that your plan is based on a profound misunderstanding, or on a failure to resist misleading evaluative appearances. And if I can empathize with you and still judge that your projects are not worthwhile, then it is not as though empathy with you will guarantee that I will be motivated to help you achieve your ends.

54 The idea of having a purpose that one regards as not worthwhile seems to border on the incoherent, but we might think, for example, of the person who doggedly pursues a dream that no longer genuinely interests them, simply because they cannot think of anything better to do with their life.
One might wonder whether my difficulty with Deigh’s account actually boils down to his deploying a different conception of empathy than the one I have been concerned with in this dissertation. Perhaps for him, empathy necessarily involves not only seeing how things look, emotionally speaking, from the other’s perspective, but also affirming that the person’s emotional perceptions are tracking the evaluative reality. To re-use some vocabulary introduced earlier, one might suppose that Deigh’s conception of empathy actually corresponds to my conception of understanding approval. If that is right, then he and I are less in disagreement, and more talking past each other. However, if that is how Deigh is conceiving of empathy, then it does becomes rather unclear what work empathy, as opposed to the mere endorsement of purposes as worthwhile, is supposed to do in banishing the psychopathic outlook. Endorsement of another’s purposes as worthwhile does not require empathy, although (as we’ve seen) empathy is often accompanied by such endorsement. The question for Deigh, then, would be: why should I think that an endorsement of another’s purposes as worthwhile is less motivationally significant if it is not a proper part of empathy?

Now let me turn back to the suggestion that empathy makes all the difference to our appreciation of others’ natures. There is a suggestion in Deigh’s account that the appreciation of others’ natures that empathy enables is bound up with the appreciation of others’ purposes as worthwhile. In fact, the reasons empathy gives me to think of you in a certain light do not depend upon appreciation of your attitudes as worthwhile, but it is true that when we empathetically grasp some normatively salient features of others’ emotions, we are bound to recognize the targets of our empathy as having certain qualities. We are bound not in the sense that we must be reflectively aware of these qualities, but rather in the sense that we cannot coherently deny them. To get at these qualities, it is helpful to think of the analogue with first-person, non-empathetic emotional experience.
When I am emotionally oriented toward the world, I am most often not thinking about my own nature. But when I do think of my own nature, I am bound to think of myself as someone who is attuned to the world’s evaluative features, even if imperfectly. After all, my attitudes are necessarily intelligible to me. I am similarly bound to conceive of myself as someone to whom things matter (an emotion is, after all, a perception of things as mattering in one way or another). Now, if I am in the grip of an emotion, reflection upon the nature of my own experience may not entail that I will conceive of myself as the architect of my life. When we feel ourselves to be the helpless playthings of our own rage or grief, the thought that we are self-determining has little resonance. If I am not in the grip of an emotion, though, I am also bound to recognize myself as someone who is able to make choices: I can accept the way things look, emotionally speaking, or I can decline to. I can act on my feelings, or I can refrain.

The self-conception to which I am bound as a feeling, thinking, and caring individual has some real motivational consequences. I need not think that the cares that grip me are genuinely worth my time. Nor, it should be stressed, does this self-conception in and of itself bar me from being cruel to myself. Indeed, I could hardly be cruel to myself in some of the ways most typical for human beings if I did not recognize myself as a creature for whom things matter, with the capacity to affirm and reject my own emotional perceptions. I could not, for instance, torture myself with the insistent thought that I was stupid to have let myself be taken in by the charm of a charlatan, or with the insidious suspicion that others care less for me than I do for them. But here is one morally bad self-directed attitude that won’t be available to me: I won’t be able to conceive of myself as unthinking, unfeeling, or uncaring. That means I won’t be able to treat myself as mere unthinking object, or an unfeeling robot, without incurring serious incoherence in my thought and action.
These facts about what is involved in inhabiting the first-person perspective in non-empathetic cases will carry over to the empathetic case, as well. If I empathize with you, I am bound, at a minimum, to recognize you as someone with a perspective for whom things matter. I must further recognize that your emotions are intelligible, rather than nonsensical. Depending upon the nature of the experience with which I am empathetically engaging, I may also be bound to recognize that you are a creature who can critically reflect upon and revise her own conception of the world. This sort of recognition is just as compatible with cruelty as was the corresponding recognition in the case of non-empathetic first-person experience. We should not underestimate the human capacity for perversity. It is possible to both recognize that a person is a thinking, caring, feeling being capable of reflection and choice and seek to make that person miserable. But the recognition empathy entails is not compatible with dismissing the target of that empathy as something less than human. I cannot both empathize with you and think that it would be appropriate to treat you as a mere thing.55

This point about how empathy shifts the ways in which we can think about treating each other might initially seem not terrifically important for the actual moral lives of people who are not psychopaths. Most of us are not, after all, ignorant of the fact that other people also have inner lives and distinct, intelligible emotional perspectives on the world. At the same time, however, it is sometimes convenient for us to neglect this fact, to pretend that others don’t have concerns, cares, or intelligible sensibilities in the same way that we do. When others feel very differently about the things we deeply love or hate, for instance, we may defensively characterize those others as lacking an intelligible emotional perspective. We might even tend to conceive of others whose interests we would rather neglect as actually feeling less fully than we do. Here I am thinking in particular of the

racist claims about black people’s inner lives that were used to justify slavery. Consider, for instance, Louisiana slaveholder George S. Sawyer’s description of “Negro character”: “the attachment to their masters, a characteristic quite common among them, is founded upon...an instinctive impulse, possessed to an even higher degree by some of the canine species.” I do not know about Sawyer in particular, but it seems likely that many who spouted this kind of vile nonsense did not genuinely believe that slaves lacked complex, human emotional lives. Rather, they found it exceptionally useful to pretend (even to themselves) that this was the case.

Empathy leaves no room for this kind of bad faith. When one merely knows that another person has an intelligible emotional outlook of their own, it is comparatively easy to set that knowledge aside for the purposes of getting what one wants, pretending to oneself that it isn’t so. When one empathizes with another, by contrast, one confronts their humanity head-on. To be actively engaged with another person’s point of view, to recognize it as like one’s own, and to nevertheless treat the other as not having a point of view in the same sense that one does oneself: perhaps it is just possible, but such a dramatic failure to act in accordance with one’s understanding will amount to a serious break-down of agency.

6. Conclusion

With that observation, I conclude my review of empathy’s contributions to the kinds of moral goods that the critics of empathy have specifically targeted. My treatment of empathy’s impact along each of these dimensions has been relatively brief. Hopefully, I have provided some reason to think that there is more to be said for empathy, in terms of its moral epistemic and moral motivation impact, than the empathy critics acknowledge. My greatest concern in this chapter, however, has been to

make the case that there is a respect in which empathy is morally good that Prinz and Bloom don’t even consider, one that is worthy of more serious attention than it has received up to this point. I have argued that the understanding empathy secures is non-instrumentally morally valuable because it satisfies a profound and widespread need. Our need to be humanely understood is complex and contextually dependent, but it is also real and serious. Not being understood is painfully isolating. This is especially true in the context of friendships, and I have suggested that empathetic understanding may in fact be necessary for true friendship.

The dimension of empathy’s goodness I have tried to bring our in this chapter may have slipped under Prinz and Bloom’s radar because they focus on empathy’s (lack of) contribution to abilities and desires whose moral value is universally acknowledged, particularly the ability to make good moral decisions and the desire to provide material aid. The good I’ve pointed to might look comparatively esoteric, for two reasons. First, despite the efforts of Collingwood and his philosophical descendants, humane understanding has received comparatively little attention from epistemologists and (especially) ethicists. And second, the very idea that a mental state of mine could be non-instrumentally morally valuable simply in virtue of another persons’ need for it to exist might seem a bit odd. Certainly we are dealing here with a very different kind of good than the material contributions to charity and efforts to save children that Bloom and Prinz discuss. However, it should be noted that humane understanding is part of a broader category of relational goods whose significance is not in the least bit obscure. We value being loved, admired, and approved of, even if we will never know about it, and even if those attitudes will never issue in further helping behavior from the one who loves and admires us. Smith claims that our need for others to harbor these sorts of feelings about us is in fact the foundation of human morality. I have not gone that far, here, but I
do think that these concerns occupy a more central place in our moral lives than many philosophers have thus far acknowledged.
CONCLUSION

To conclude this dissertation, I will first briefly review the main lines of argument I have pursued. Then, I will consider how the analysis of empathy's nature and significance I offer in the latter two chapters of the dissertation measures up to the standards and expectations I introduced in Chapter One. In the course of offering that assessment, I will also indicate some possible directions for future research.

1. The course of the investigation

We began with a characterization of empathy that I take to reflect powerful and widely shared intuitions, the core of which is: empathy is (1) an imagination-implicating mental activity that (2) involves getting at others’ experiences “from the inside” and that (3) somehow involves the emotional life of the empathizer. I then sketched out an enduringly popular, broad vision of how empathy matters, the bridging picture. According to this picture, our most important forms of self-concern and self-knowledge are stronger, more durable, closer to inevitable, and more immediate than our corresponding knowledge of and concern for others. Our primitive position (the one we are born into, or the one we most naturally occupy) is doubly (epistemically and motivationally) egocentric. Happily, though, the story goes, empathy offers a or the means of escape from the solipsism and egoism that would otherwise be our natural state, by taking something of our way of relating to ourselves and what is ours, and transferring it to our relations with other people and what is theirs. The question for me was whether this bridging picture could be developed in a coherent and realistic way, one that distorts neither the nature of empathy nor the nature of our other-oriented concern and understanding.

In search of an answer, I first turned to the earliest extended philosophical description of how
empathy matters, in Hume’s *Treatise*. Hume proposes that the transference of mental vivacity characteristic of empathy moves us first from total ignorance about other’s mental states to knowledge of them, and then from total indifference to others’ suffering to concern for that suffering that matches the others’ concern. The ideas advanced in the *Treatise* about empathy’s epistemic and moral significance remain popular. They figure centrally in the contemporary revival of interest in empathy. However, I argued that proposals in the Humean vein fail to deliver a viable version of the bridging picture. His account, and others like it, end up making it look as though empathy simply contributes to the expansion of egoism, and/or as though empathy depends for its motivational efficacy on a kind of epistemic mistake. Some Humean accounts, we saw, do avoid making empathy’s functions hinge upon epistemic error, but at the cost of casting empathy as an idle wheel, rather than a critically important bridge between oneself and others.

Strikingly, Humean articulations of the bridging picture sit uneasily with two widespread intuitions about empathy, namely (1) that empathy’s *feeling* nature matters to the quality of the knowledge it affords, and (2) that the way empathy matters to our care for others has everything to do with the way it changes our understanding of their inner lives. Taking my cue from those latter two thoughts, I developed an alternative account of what empathy can do for us, which is summed up by the *humane understanding thesis*: empathy furnishes a distinct epistemic good, humane understanding, which is necessary for (and, indeed, partially constitutive of) fully virtuous relations with other people. Some hints of this sort of proposal are present in Adam Smith’s *The Theory of Moral Sentiments*, so I relied on that work, as well as some insights from contemporary theorizing about the emotions, for support.
Adam Smith holds that empathy is the unique means by which we judge the propriety of other’s emotions. I drew out the kernel of truth in this surprising claim. Emotions inevitably present the world in a certain evaluative light. When I see my situation as calling out for my anger, or fear, or adoration, my emotion appears to me to be properly responsive to my condition. I can judge my emotion to be inappropriate, all things considered. However, I cannot both be angry at your betrayal and be totally mystified by my own anger, or resent you but see no absolutely nothing that appears to call out for resentment. This insight applies to empathy. When we empathize, we feel genuine (thought-directed) emotions. And when we feel these emotions, we cannot fail to appreciate the intelligibility of the other’s original emotion first-hand. This appreciation constitutes humane understanding of the other’s emotion. I further argued that humane understanding is not available through mental activities that do not involve perspective-taking, nor is it available through forms of perspective-taking that do not involve emotional engagement.

Recognizing that empathy is the unique means by which we secure humane understanding of others’ emotional perceptions (and, by extension, the choices and actions that stem from those perceptions) enables us to appreciate aspects of empathy’s moral significance that its most prominent critics have neglected. In empathizing, we can be in fellowship with others and relieve their very real distress at not being understood. We do not always want to be humanely understood by all others. Our need for humane understanding is complex, subject to the interaction of multiple contextual factors. In the absence of others’ humane understanding, though, we are psychologically isolated in a way that can be very painful indeed. After highlighting empathy’s value as a cure for this isolation, I went on to argue that empathy also plays an important role in the development of both our moral beliefs and our emotional sensibilities, and I concluded with the observation that empathy prohibits us from
seeing others in certain morally bad ways. In empathizing, we directly confront others’ humanity, and this makes it very difficult for us to treat others as unthinking, unfeeling things.

2. How my account measures up

In Chapter One, I raised the question of whether the bridging picture could be vindicated, and I introduced four desiderata for an account of empathy’s nature and significance. Now that I have laid out my view of what empathy is and how it matters, we can revisit these points, and consider (1) whether my view has ended up preserving the bridging picture, and (2) whether my account satisfies the desiderata I proposed.

2.1. The fate of the bridging picture

In the end, my account does not amount to a wholesale rejection of the bridging picture. I have not attempted to argue that our self-understanding and self-care is in no respects privileged (which would amount to denying that there is some gap between ourselves and other people for empathy to bridge). However, some aspects of my account of the bridging work that empathy can do actually hinge upon the denial that our primitive state is one of extreme solipsism and egoism. To bring how this is so, consider first how my view differs from the traditional thought about empathy’s impact on our concern for others. Because it is often an effortful, even difficult enterprise, rather than an automatic process, the pursuit of empathy is typically something we have to have reason to undertake. Why do we empathize? Well, we might have several different reasons. Sometimes, admittedly, our empathetic efforts are driven by concerns that have little to do with benevolent interest. We might set out to empathize because (as both Hume and Smith remark) it is pleasing to vividly imagine being in exciting or luxurious circumstances. Or we might even be driven by cruelty, as in the case of a torturer who wants to savor the pain of his victim from the victim’s point of view.
But we would not be moved to empathize by others’ pain at not being understood unless we already had an orientation of benevolent concern toward the other. That is to say, at least sometimes, we empathize because we care, rather than caring because we empathize.

Now consider how my view differs from the traditional thought about the epistemic bridge that empathy provides between self and other. I have argued that the first-person perspective is special because when one experiences an emotion from the inside, one is bound to recognize the emotion’s intelligibility. Through empathy, we can imaginatively engage with others’ emotional perspectives, and thereby come to humanely understand them. In Chapter Four, I pointed out that pursuing humane understanding of others’ evaluative perceptions can bring us closer to others in this respect (among others): when another person’s emotional response to their situation initially strikes us as strange or unaccountable, but we trust their testimony that their situation is exciting, or terrible, or frightening, we can use that testimony to guide our efforts to extend our own emotional sensibilities, such that we apprehend the presence of those evaluative properties for ourselves. When our emotional sensibilities shift in this way, they come to align more closely with those of the other. This evolution in our sensibility would not, however, be possible if our knowledge of others’ inner lives arose solely through empathy. We first need to learn that others’ outlooks do not accord with our own before we can use their alternative outlooks to orient our efforts at extending our sensibilities, and the discovery that our outlooks do not coincide cannot, ex hypothesi, be achieved through empathy.

This story about the role of testimony brings up another respect in which my version of the bridge empathy forms differs from the traditional bridging picture. On the traditional conception, the bridge is one-way. Empathy allows us to exploit our immediate and robust grasp of our own inner life to learn about others’ inner lives, which would otherwise remain mysteries to us, but the picture
does not make room for empathy to teach us new things about ourselves. On my view, our understanding of our own inner lives is itself affected by our empathetic efforts, at least in the following sense: we learn more about ourselves in empathizing when we discover that we actually can make sense of attitudes that we might have initially assumed would be hopelessly foreign to us, or when we find that attitudes we endorsed at correct do not actually ring true for us, emotionally speaking. Empathy may also contribute to self-understanding in the sense that other people's efforts to get where we are coming from can change our understanding of our own feelings. As I mentioned in my discussion of the inarticulacy of emotions, our friends' efforts to see how things are for us from our point of view can help us to identify more precisely what exactly our emotions are about.

In this dissertation, I have really only touched upon empathetic efforts' influence on our self-conception. There are intriguing suggestions in Smith's *Theory of Moral Sentiments* that give empathy a role to play in our self-understanding even more significant than the contributions I've just mentioned. For Smith, our ability to empathize is in fact essential for the reflexive awareness of our own inner lives. He writes:

> To a man who from his birth was a stranger to society, the objects of his passions, the external bodies which either pleased or hurt him, would occupy his whole attention. The passions themselves, the desires or aversions, the joys or sorrows, which those objects excited, though of all things the most immediately present to him, could scarce ever be the objects of his thoughts. The idea of them could never interest him so much as to call upon his attentive consideration. (TMS 110)

Smith holds that when we come into society, and become aware that others have their own views of the world and of us, we begin to form a conception of ourselves as someone whose vision of the world could be approved of as accurate or dismissed as false. It is through empathizing with others and their perceptions of us that we acquire enough distance from our own attitudes and perceptions to be able to see them as attitudes and perceptions.
Is empathy really necessary for self-awareness in the way that Smith proposes? Should we perhaps go even further, and say with Samuel Fleischacker that “we can maintain a perspective only if we can engage in Smithian empathy”?\(^1\) The true extent of empathy’s impact on our self-understanding is one of those questions that demands further investigation. I will not pursue it any further here, however. Instead, I will turn back to the desiderata I introduced in Chapter One.

2.2. Revisiting the desiderata

The first hope I introduced for an account of empathy is that it would be sensitive to the ways in we actually value empathy in practice. In recent debates, both critics and defenders of empathy have tended to focus on empathy’s contribution to the production of altruistic desire. This is not surprising. After all, there is a long history of insisting that empathy drags us up out of egoism, and the moral value of altruistic desire is something that ethicists from diverse traditions can all readily acknowledge. However, when we look at how people actually talk about their need for empathy, the prospect of ginning up altruistic motivation in the empathizer hardly seems to be the first thing of their minds. In self-help guides, internet commenting-boards, and other non-academic fora, when people touch on empathy’s importance to them, the value of such imaginative engagement with other perspectives is frequently expressed in terms of its contribution to knowledge or understanding. So, we find a theatre-goer and caretaker professing disappointment that a stage production couldn’t “imagine what it was like to be the parent of a disabled child,” and a Michigan prisoner and artist making this request: “I would ask that those who judge us to perhaps look past the blue and orange state clothes we wear, and to try to practice empathy. Please try to understand us.”\(^2\) The proposal that empathy is non-instrumentally valuable because it uniquely satisfies the need

\(^1\) Fleischacker (unpublished MS).

\(^2\) Birkett (2015); Allen (2018).
to be understood aligns well with the ways we talk and think about empathy outside of a philosophical context.

The second desideratum for an adequate account of empathy was that it would show how each of empathy’s key features matters to its significance. In Chapter Two, I explicitly argued that both perspective taking and emotional engagement are necessary for humane understanding. I have not done the same for the third key feature of empathy, its imaginative nature. As it happens, the claim that empathy’s imaginative nature makes any difference to its epistemic significance has also been challenged. Kendall Walton argues that just the same epistemic goods can be secured by actually, rather than imaginatively, inhabiting the position of another. That challenge is less interesting than the ones directed at the significance of perspective taking and emotional engagement, though. It may be that if we could actually inhabit the precise situations of the many diverse others we empathize with, then imagination would have no role to play in humane understanding. But it is practically inconceivable that even in cases where we actually do inhabit a situation very similar to another’s (we are both in the middle of the same wild rafting trip, we have both suffered the loss of our child) there will be no differences in our situation such that an imaginative adjustment to accommodate these differences could yield a finer, more accurate empathy (we are both in the boat, but you are in the bow, and are smaller than me; we have both lost our child, but you gave birth to him, and so forth).

The third desideratum was that the account would explain the relationship between empathy’s moral and epistemic significance. This has been the principal task of the latter two chapters. I have argued that, contrary to what Hume’s account suggests, empathy’s moral and epistemic significance are
deeply intertwined. Empathy secures a form of understanding that the recipients of empathy value for its own sake. Humane understanding is simultaneously a moral good and an epistemic one.

A reason to think I have not yet adequately explained the relationship between empathy’s moral and epistemic significance, despite my dogged efforts, is that I have not yet fully settled the question of whether (or when) it is morally good to try and appreciate the intelligibility of morally bad perspectives. I have said that we have profound needs to be made sense of, needs which others can have moral reason to try and meet. But often, the outlooks that we desperately wish others would empathize with are not altogether salubrious or morally pure. We wish others to empathize with our annoyance at slights, or our desire for revenge, or our envy at our rival’s successes. One strain of thought in the virtue ethical tradition holds that that there are some things it is better not to know, if one wants to be greatly good. Might we also think that there are some things it is better not to appreciate, and that the intelligibility of vicious perspectives is one of those things?

According to one sort of characterization of virtuous vision, morally bad possibilities do not even show up for the virtuous person as attractive, and bad ways of achieving ends are, as Bernard Williams’ puts it, “unthinkable,” not even salient. So, for instance, a saint will not be able to see petty slights as important, and will recoil at the thought that anything in the world could call out for envy or vengeance. If bad ends are unthinkable for the truly virtuous, then perhaps the virtuous will be limited in their empathetic abilities in ways that more vicious (or at least akratic) types will not be, which means they will not be especially well positioned to empathize. We need to ask: will improving ourselves, morally, mean that we will no longer be able to make sense of attitudes that we

---

3 Driver (2004) is a particularly good example of this style of analysis.

4 Williams (1981): 129.
used to “get”? And if so, does this mean that only imperfectly virtuous people will be able to meet others’ genuine needs for empathetic understanding?

I have suggested that when it comes to moderately vicious perspectives, doing right by others may sometimes require us to empathize. I am less confident that the same goes for deeply wicked perspectives. Whether or not it does, in the end, further reflection on the question of when and why we ought to resist empathy with morally bad outlooks will have implications for the broader ongoing conversation about the nature of virtuous vision.

Finally, we come to the forth desideratum, namely that it would be good for an account of empathy to identify the root causes of any significant distortions in rival conceptions of empathy. Over the course of this dissertation, I have pointed to a number of sources of difficulty for both historical and contemporary accounts of empathy. Let me briefly review some of the most important of these.

One source of difficulty is that empathy involves emotion, and emotion is a component of humans’ inner lives that has tended to give philosophers a lot of trouble. They have intentional content, but they also involve bodily responses. They move us to action, but they are not the same thing as mere urges. Faced with the jumble of characteristics associated with emotion, philosophers have tended to go in one of two directions. One can describe an emotion as a chimera, an assemblage of different elements. This is, roughly, Jane Heal’s strategy: for her, emotions are made up of (1) a judgment and (2) non-intentional elements such as sensation. Or, alternatively, one can winnow empathy down to just one of its apparent elements. Smith, for instance, seems tempted at times to cast emotions as judgments.
Depending upon which of these interpretive courses one takes, the downstream problems for account of empathy’s significance will be somewhat different. If one conceives of emotions as judgments, then it will not be clear what benefits empathy can secure than cannot be achieved through unemotional judgment. And if one conceives of emotions as a combination of a judgment and non-intentional elements, then it will not be clear what of epistemic significance can be secured through empathy, above and beyond what the judgmental element of an empathetic emotion secures. Consequently, accounts of empathy’s significance have not tended to do justice to the intuition that empathy’s emotional nature makes all the difference to the knowledge or understanding of others that empathy affords. The solution to this problem is to reject both chimerical and reductive characterizations of emotion in favor of the view that emotions are evaluative perceptions.

A second problem has been the tendency to neglect the point of view of empathy’s recipients. Recently, both critics and supporters of empathy have tended to focus on what empathy means for the production of “pro-social behavior.” However, I have argued that this sort of focus obscures what is so special about empathy, from the perspective of those who need it. We don’t just want others to empathize with us so that they will be more likely to help us. Rather, we want to be understood, because not being understood is painfully isolating. I think the reason this perspective has not been taken seriously goes all the way back to the historical origins of the bridging picture. The Early Modern philosophers who developed the bridging picture of empathy’s significance were motivated by the need to respond to Mandeville and Hobbes’ claims that what look like altruistic behaviors are in fact driven by egoistic impulses. Against the backdrop of these scandalous claims, a defense and explanation of human altruism was naturally regarded as an urgent priority. Some contemporary work on empathy seems to have inherited this some orientation. Explaining how we can possibly come to be moved by others’ plights continues to be seen as more important than
exploring other aspects of empathy’s moral significance. Consequently, some of empathy’s most interesting and important contributions to our moral lives, such as its role in friendship, and its contribution to the development of our emotional sensibilities, tend to be obscured.

Finally, we have seen that empathy’s principal strengths lie in constituting and supporting the goods associated with fellowship and intimacy. It bears more vexed relation to goods closely tied to the moral pole of respect, which is associated with standing at a distance, and with recognizing and honoring difference. As we have seen, others’ attempts at empathy may be experienced as invasive, because empathy involves accessing one’s worldview from the inside, or as detrimental to the aim of being regarded as a full equal, because of the defeasible but real connection between empathy and pity. This fact about empathy’s particular strengths and weaknesses helps to explain why philosophers have been tempted by the thought that empathy’s moral significance derives from its dissolution of the barriers between self and other: the former view is a natural (if highly distorting) exaggeration of a truth about empathy’s powers.

3. What remains to be said

I have already indicated some of the hard questions about empathy that are still waiting on full and proper answers. More needs to be said about the moral status of empathy with bad perspectives. There is also plenty of room for a further story about the respects in which our empathy with others shapes ourselves and our self-conception. The description of empathy’s character and epistemic import that I have given here will also provide a solid base for an exploration of how empathy relates to morally important attitudes distinct from humane understanding. I have already begun to do this exploratory work with regard to the connection between empathy and trust.5 We should also

5 Bailey (forthcoming).
investigate the connection between empathy and forgiveness, since our ability to forgive is likely related to our ability to humanely understand the bad attitudes or actions of others. Empathy’s relation to the virtue of humility is likewise ripe for exploration, since we tend to think of empathy as an antidote to the vice of being judgmental, and a tendency to be judgmental is intimately connected with an arrogant sense of superiority. All of these further avenues of inquiry will have to wait for another day, however.


Hollan, Douglas. “Empathy and Morality in Ethnographic Perspective.” In *Empathy and Morality,


Persson, Ingmar, and Julian Savulescu. “The Moral Importance of Reflective Empathy.” *Neuroethics*


