The Role of Empathy in Moral Inquiry

by

William G. Kidder

A Dissertation

Submitted to the University at Albany, State University of New York

In Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

College of Arts & Sciences

Department of Philosophy

2021

ABSTRACT

In this dissertation, I defend the view that, despite empathy's susceptibility to problematic biases, we can and should cultivate empathy to aid our understanding of our own values and the values of others. I argue that empathy allows us to critically examine and potentially revise our values by considering concrete moral problems and our own moral views from the perspective of another person. Appropriately calibrated empathy helps us achieve a critical distance from our own moral perspective and is thus tied to impartiality in moral inquiry. In defending this role for empathy in moral inquiry, I draw on empirical work from psychology and neuroscience to support a constructionist account of emotion, according to which we can develop more wide-ranging, fine-grained emotion concepts and empathetic capacities by seeking diverse experiences, communication, and engagement with art. I then defend the value of this effortful correction of empathy bias, arguing (1) that impartial moral inquiry ought to utilize empathy as a check on motivated reasoning and presumptions regarding what count as appropriate solutions to moral problems, and (2) that compassionate moral inquiry ought to involve empathy as a means of recognizing others as authentic moral agents that can make valuable contributions to moral debate. Lastly, I draw on insights from pragmatist philosophy to critique Adam Smith's empathy-based account of the "impartial spectator" and defend a conception of impartiality grounded in fallibilistic, empathetic method.

Keywords: empathy, moral inquiry, impartiality, emotion, Adam Smith, pragmatism

ACKNOWLEDGMENTS

I would like to thank my advisor, Jason D'Cruz, and the other members of my dissertation committee, Brad Armour-Garb, Brendan Gaesser, and Ariel Zylberman, for their support and helpful feedback throughout the process of writing this dissertation. I would also like to thank my daughter, Anna, whose perspective is a source of joy and inspiration.

TABLE OF CONTENTS

Introduction

> Were it possible that a human creature could grow up to manhood in some solitary place, without any communication with his own species, he could no more think of his own character, of the propriety or demerit of his own sentiments and conduct, of the beauty or deformity of his own mind, than of the beauty or deformity of his own face. All these are objects which he cannot easily see, which naturally he does not look at, and with regard to which he is provided with no mirror which can present them to his view. Bring him into society, and he is immediately provided with the mirror which he wanted before.
>
> -Adam Smith, *The Theory of Moral Sentiments*

The focus of this dissertation is empathy, by which I mean the capacity to inhabit, to some degree, the perspective of another, including the other's emotional perspective. I begin with Smith's observation because it brings to light two central themes that I will address. The first is the difficulty in recognizing the "propriety and demerits" of one's own sentiments and conduct, the difficulty in engaging in unbiased, self-critical moral reflection. Contemporary research has done much to corroborate Smith's claim that the moral propriety of our own conduct and views is often not something that we are inclined to critically inspect. Part of my aim is to identify how this problem manifests itself both when our empathetic engagement is biased and when we fail to empathize entirely.

Empathy can lead us to favor members of a perceived in-group at the expense of members of perceived out-groups, to favor individual concerns at the expense of the concerns of larger groups, and to favor the concerns of those who are more proximate at the expense of those who are more distant. Yet, I will argue that despite this susceptibility to bias, we can and should make an effort to utilize empathy in moral inquiry. This effort expands our moral perspective; it is an effort to maintain openness to evidence that allows us to engage in a less partial assessment of our own moral beliefs and conduct. It is true that empathy bias presents a significant obstacle to this assessment insofar as it hinders one's ability to take on moral perspectives that differ from

one's own to the extent required to provide the sort of "mirror" through which one's moral beliefs and conduct can be accurately reflected. But fortunately, as Smith concludes, we can in fact find such a mirror in society. This is the second central theme of the dissertation. I will argue that our ability to critically evaluate and revise our own moral beliefs and conduct is tied to our ability to empathize with a wide variety of perspectives; these perspectives serve as important evidence in moral inquiry. Thus, we should make the effort to develop a wide-ranging and fine-grained capacity for empathy. I do not deny that our inherent susceptibility to empathy bias presents a serious obstacle to cultivating this ability, but my aim here is to identify and defend a method that allows us to correct for empathy bias and utilize empathy in moral inquiry, rather than tamp down empathy and fail to realize its potential contributions to the pursuit of impartiality.

While the role of empathy in morality has been debated amongst philosophers in the Western[1] tradition at least since David Hume (1739/2000) and Adam Smith (1759/1982) made "sympathy"[2] a central component of their ethical systems in the 18th century, the topic has received renewed interest in recent years as psychologists and philosophers, in particular the psychologist Paul Bloom (2016) and the philosopher Jesse Prinz (2011a, 2011b), have focused on the problem of empathy bias and argued that it is morally problematic. Critics have suggested that empathy's susceptibility to bias is significant enough to discourage us from relying on empathy in our moral lives. I will argue that this conclusion is too strong. While I recognize that empathy is susceptible to problematic biases, I argue that these biases are correctable and are

1 The concept of adopting the perspectives of others plays a prominent role in the Confucian tradition. Mengzi argued for the importance of empathy as a motivator for altruistic behavior. See Slote (2010b).
2 While Smith and Hume use the term 'sympathy', they refer to the capacity to take on the perspective of another person. This capacity has generally come to be known as empathy and is the sort of capacity that I will be discussing here. I take 'sympathy' to refer to a concern with others that does not necessarily require a sharing of the other's perspective.

worth correcting. Empathy bias is worth correcting because empathy can and should play a valuable role in providing evidence in the process of moral inquiry.

Arguments that seek to establish that empathy is susceptible to morally problematic biases appeal to a convincing array of empirical evidence.[3] As noted above, this includes bias towards empathizing with members of perceived in-groups, rather than with members of perceived out-groups, and bias towards empathizing with those who are more proximate rather than with those who are geographically distant but in more need of aid.[4] In addition, as Bloom emphasizes, empathy is problematically innumerate in that it can lead us to focus on the experiences of individuals at the expense of prioritizing the suffering of large groups of people.[5] Thus, argue Bloom, Prinz and other critics of empathy, if empathy is driving our moral judgments, then we will be more apt to favor agreeing with and helping the few individuals with whom we are especially emotionally connected and will do so as a result of morally irrelevant social and geographical circumstances, leading us to ignore or downplay the concerns of vast numbers of people who may be in more need of consideration and aid.

Unfortunately for moral inquiry, this bias may also render us less able to realize that we are in fact biased, as we will tend to empathize with those who view moral problems from perspectives that fit with our own and will take this empathetic experience as evidence in support

3 It should be noted, however, that there is also a significant amount of evidence linking empathy to motivating altruistic behavior. For surveys of this sort of evidence, see de Waal (2009), Rifkin (2009), and Batson (2011). While Bloom and Prinz recognize this evidence, their claim is that the "dark side" of empathy, to use Bloom's language, outweighs its potential for motivating altruistic behavior, which can be motivated via other means, e.g., via other moral emotions such as outrage, guilt, etc. for Prinz, and via "rational compassion" for Bloom. The role that I defend for empathy is not one of motivating altruistic behavior but rather is one of aiding the critical evaluation of moral motivations, so I will not appeal to evidence that stresses empathy's potential to motivate altruistic behavior. My argument stresses that altruistic, compassionate motivations ought to motivate a certain kind of empathetic effort.

4 See, for example, Brown et al. (2006), Batson and Ahmad (2009), Xu et al. (2009), Mathur et al. (2010), Hein et al. (2010), Cikara, et al. (2011a), Cikara and Fiske (2011), and Cikara et al. (2014). This empirical work on empathy bias will be discussed at length in Chapter 2.

5 See, for example, Batson et al. (1995) and Kogut and Ritov (2005). Again, this empirical work will be discussed at length in Chapter 2.

of the idea that we have the correct perspective on the moral problem in question. This is a dangerous feedback loop in which we end up with both myopic moral perspectives and complacency regarding the need to broaden those perspectives.

Evidence of empathy bias is worth taking seriously and needs to be addressed when defending any moral philosophy that stresses the importance of empathizing. It is not my aim here to deny the significance of this evidence, nor to deny that empathy can operate problematically in the moral realm. On the contrary, I believe that recognizing the prevalence and strength of empathy bias is critically important in cultivating a healthy moral life.[6] My aim is to outline an approach that can remedy the problematic aspects of empathy and retain the benefits of employing empathy as a tool for gathering evidence that is beneficial to critical moral inquiry.

I will first clarify the conception of empathy that I take to be morally important and worth remedying. This is the focus of Chapter 1. 'Empathy' is defined in a variety of ways across work in psychology and philosophy. These definitions span a wide spectrum of emotional and cognitive capacities and behaviors, so it is important to precisely define the phenomenon I have in mind when I claim that empathy is morally important and is worth cultivating.[7] My goal in Chapter 1 is to articulate necessary and sufficient conditions for empathy that include interrelated affective and cognitive components, and to highlight the role of empathy in our understanding of

6 There is evidence suggesting that simply believing that it is possible to change one's capacity for empathy can cause one to become more empathetic and to make long-term efforts to improve one's capacity for empathy. See, for example, Schumann et al. (2014), which draws on Carol Dweck's research on the role that "mindsets," people's beliefs about their own psychology, play in altering behavior. This work is summarized in Dweck (2006). It is important to keep this idea in mind when drawing attention to empathy bias, as the goal is not to point out an inalterable human deficiency, but to bring a problematic susceptibility for bias to light so as to pursue the necessary behaviors to correct the bias. Research on mindsets and empathy highlights how important it is to approach the problem of empathy bias with a belief that empathetic capacity is malleable. Such a belief is not mere wishful thinking. I will defend empathy's malleability in Chapter 3. Ultimately, I argue that an awareness of empathy's malleability, coupled with the fallibilist mindset that I defend in Chapter 6, should lead one to pursue the behaviors necessary to correct empathy bias, behaviors that I discuss at length in Chapter 3.

7 For a discussion of the range of definitions of empathy, see Batson (2009).

the emotional perspective of others. Some define 'empathy' merely in terms of the ability to understand the perspective of another person, with no necessary affective component, but I take empathy to involve a simulation of another's affective experience. Empathy conceptualized without a necessary affective component is often referred to as cognitive empathy. This conceptualization of empathy is consistent with those who defend a theory-theory view of our ability to understand other minds (i.e., our "theory of mind" or "ToM"). Proponents of theory-theory argue that we do not simulate the perspectives of others, but rather theorize about others' mental states based on our understanding of folk psychological laws. I will challenge this theory-theory approach and argue that our understanding of others' mental states, particularly emotional states, necessarily involves simulational empathy.

Cognitive empathy and theory-theory are not the focus of the sort of critiques advanced by Bloom and Prinz. The primary target of their criticisms is the experience of empathetically simulating another's emotion.[8] The problematic biases discussed above are often couched as emotional biases that run counter to a more careful, deliberative approach grounded in reason. In this vein, Bloom argues that empathy should be replaced by "rational compassion" that can avoid such emotional biases. Thus, one could perhaps sidestep the sort of criticisms highlighted by Bloom by advocating for mere cognitive empathy. One could argue that the only sort of empathy worth defending is the sort that is defined as mere other-oriented perspective taking, absent any shared feeling with another. But this is not my goal. I defend empathy as a more robust phenomenon, involving both a necessary cognitive component *and* a necessary affective component. I take the affective and cognitive components of empathy to be crucially connected

8 This point is explicitly stated by Bloom (2016, pp. 35-39).

in the simulation of emotion and take this connection to be key to the effortful correction of empathy biases in moral inquiry.[9]

Given that I take empathy to be necessarily affective, my account needs to answer criticisms that emphasize the emotional nature of empathy bias. I discuss a variety of such criticisms at length in Chapter 2, highlighting some of the empirical work that has drawn attention to the empathy biases discussed above. There are two overarching questions that the mass of evidence of empathy bias forces us to consider: (1) Can we correct empathy bias? (2) Should we correct empathy bias? That is, is it worth it to correct the problem rather than seek other, potentially more effective approaches to morality that do not involve empathy? Answering these questions in the affirmative will be the focus of the remainder of the dissertation.

I begin by defending a method of effortful correction of empathy bias in Chapter 3. My account of how this is possible relies on understanding empathy in terms of the conceptual act theory of emotion, a constructionist account defended by Lisa Feldman Barrett (2005, 2011, 2015, 2018).[10]

As discussed above, much of the concern regarding empathy bias occurs because it is couched as an *inability* to appropriately consider certain perspectives due to reasons that are irrelevant or counterproductive to addressing the moral problem at hand. However, this framing of the problem is based on a conceptualization of empathy as an involuntary process involving mere affective matching, rather than as a process necessarily involving an experience of affective matching that is linked to one's effort to understand the perspective of another. Once empathy is

9 I will explore this connection in terms of constructionist theories of emotion, but I take my approach to be generally in line with work in neuroscience and psychology that rejects the neat, dichotomous distinction between emotion and cognition. For a general critique of approaches in psychology that emphasize such dichotomies, see Melnikoff and Bargh (2018). For work on the interrelation of emotion and cognition specifically, see, for example, Ashby et al. (1999), Dalgeish and Power (1999), Forgas (2001), Lerner et al. (2004), Phelps (2005), and Gray et al. (2005).

10 See also Barrett et al. (2015).

conceptualized as outlined in Chapter 1, we can see that the interrelation of other-oriented perspective taking and affective matching allows us to frame the problem of empathy bias not as a problem of involuntary bias, but rather as a problem of bias that can be corrected through one's effort to engage in more nuanced and diverse experiences of other-oriented perspective taking.

Barrett's conceptual act theory provides an explanation for how effortful other-oriented perspective taking can combat empathy bias. Roughly, Barrett argues that emotions are not biologically hard-wired reactions to stimuli that all human beings share, but rather are constructed according to conceptual schemas that are built up from experience within the particular cultural and linguistic environment in which an individual develops. This is an anti-essentialist account in that emotions are not inborn, uniform capacities possessed by all human beings and marked by identifying physiological fingerprints. Rather, emotions are unique experiences constructed by the individual who experiences them according to that individual's experiential background. This is a top-down process: instances of an emotion are constructed via predictive coding in the brain, and this coding is the result of an individual's prior experience and knowledge. For example, an individual who experiences a fear of heights will experience this fear not because heights trigger an innate fear circuit in the body, but because that individual has built up a concept of fear that incorporates certain information about heights based on prior experiences with heights (e.g., the painful experience of falling from a tree as a child). Prior experiences have wired that individual to predict certain physiological changes and sensory experiences associated with fear in the contexts of the environment of heights. By contrast, an individual who lacks a fear of heights will have a different conceptual schema based on different experiences with heights, and thus will not predict the same sorts of physiological changes or sensory experiences when standing on a balcony, etc. On Barrett's model, the brain is

continuously constructing predictions of possible experiences based on prior experience, and these predictions are modulated by outside stimuli. Emotions are a particular instance of this process of prediction, identified not by hard-wired neural networks necessarily involved in the construction of specific emotions, but rather by the context in which a particular prediction of experience occurs and by the particular concept, built up over the course of an individual's experience, underlying the construction.

There are no necessary and sufficient conditions to define an emotion across all individuals, as each individual's unique conceptual schema leads her or him to construct emotions differently as a result of her or his prior experience. In terms of empathy, this anti-essentialist, constructionist account is significant in that, on such an account, the best way to understand the diversity of others' emotional experiences is to understand the diverse conceptual schemas and cultural contexts that underlie them. Understanding these conceptual schemas is the goal of effortful other-oriented perspective taking. Making the effort to understand the conceptual schemas involved in another's emotional experience puts us in a better position to engage in the top-down construction required to *experience* that individual's particular emotion. Emotional experience is constructed in terms of predictions based on prior knowledge and experience, so we will be better able to empathetically construct the emotions of another if we make the effort to understand her or his prior knowledge and experience. The constructionist theory of emotion allows us to see how we can become better empathizers by making the conscious effort to expand our own experience.

I argue that there are three general routes to this effortful expansion of experience. First, one may pursue direct engagement with novel experiences so as to subject one's own emotion concepts to refinement based on the demands of these experiences. I call this the embedded

approach. The goal is to develop what Barrett calls "emotional granularity" in one's emotion concepts by subjecting those concepts to the crucible of lived experience. In doing so one cures areas of experiential blindness to certain environments and contexts, environments and contexts that may be involved in the emotion concepts of others. Pursuing diverse experiences is a means of increasing one's potential for conceptual overlap between one's own emotion concepts and the emotion concepts of others. Such overlap enables one to combat empathy bias by leading one to draw on more fine-grained and diversely informed emotion concepts when attempting to construct the emotional experience of others.

Second, one may pursue direct communication with those whose experiences differ from one's own. I call this the communicative approach. While the embedded approach enables one to refine one's emotion concepts via directly confronting novel environments, the communicative approach pursues the same goal indirectly via effortful engagement with first-hand accounts from those who have had direct experiences that one may not have had oneself. This approach is especially helpful given that many morally relevant experiences are either impossible to directly experience (e.g., a man directly experiencing the sex-based discrimination experienced by women, or a white person experiencing the racism experienced by Black people in the United States), or are imprudent to pursue (e.g., the experience of long-term homelessness or drug addiction). While bias may make it more difficult to empathize with those whose experiential background is vastly different from our own, this bias need not prevent us from communicating with such individuals, and such communication has the potential to reduce empathy bias insofar as it has an impact on our emotion concepts.[11] Communicating with those who have different experiential backgrounds is an indirect means of exposing our own emotion concepts to potential revision based on the experiences of others and thus to increasing the potential for conceptual

11 The empirical evidence in support of this claim will be discussed at length in Chapter 3.

overlap between our emotions and the emotions of those from different backgrounds. In other words, communication, like direct experience, is capable of curing areas of experiential blindness and developing emotional granularity. This granularity in turn enables us to construct the emotions of others with more accuracy—that is, it enables us to become better empathizers.

Third, effortful engagement with art, particularly narrative art, is a means of understanding perspectives that differ from one's own perspective. I call this the imaginative approach. This idea is consistent with empirical work suggesting that engagement with literature and television dramas can play a role in improving empathy,[12] and it draws on the idea that art is especially suited to express the sort of nuanced perspectives and contexts involved in developing more wide-ranging, fine-grained emotion concepts. While we should pursue interactions with those who have different perspectives in our everyday lives through the embedded and communicative approaches, engaging with these perspectives through art provides a particularly powerful experimental space[13] in which to try on the perspectives of others and challenge our moral beliefs. It is difficult to engage with the range of relevant moral perspectives in our everyday lives; it is not a realistic option for most of us to continuously travel the world, actively seeking out interaction with as many diverse perspectives as we can in order to challenge our beliefs through empathizing. There are practical considerations that constrain efforts to empathize. Engagement with art can allow one to sidestep some of these practical constraints on the cultivation of unbiased empathy. While I may not be able to travel to Japan or Afghanistan to engage with the perspectives of Japanese or Afghani people, I can engage with Japanese or Afghani literature, films, paintings, photography, etc. One can visit libraries and museums, and it

12 For example, see Black and Barnes (2015) and Vezzali et al. (2015)
13 There is also a sense in which art provides a particularly *safe* space to conduct moral inquiry. Keen (2007) argues that the novel's status as fictional increases the reader's empathetic response by reducing common guarded reactions to real others.

is a benefit of the globalized, technology-saturated world that we live in that literature, music, films, photography, and paintings from all over the world are immediately available to millions of people via the Internet.

My argument for the imaginative approach draws on the work of Martha Nussbaum (1983, 1985)[14] and is in line with views expressed by Gregory Currie (1995), who sums up the value of fictional narratives thusly:

> [I]t is often hard for us to sustain an imaginative exploration of a complex situation. That is where fiction comes in. Fictions can act as aids to the imagination – holding our attention, making a situation vivid for us, and generally drawing us along in the wake of the narrative. If they can help us enter empathetically into the characters, we can come to feel what it is like to be those characters, make their choices, pursue their goals, and reap the rewards and the costs of their actions (pp. 163-164).

Literature can express a pace and depth of thought and feeling that is difficult to express in everyday interactions. Understood in terms of Barrett's conceptual act theory, my claim is that narrative art can express fine-grained emotion concepts held by the characters involved and that empathetically[15] engaging with this level of emotional granularity in a fictional context allows the reader to reflect on and refine one's own emotion concepts such that one is better suited to construct the emotions of others in the process of empathizing outside of fictional contexts.

The embedded, communicative, and imaginative approaches to correcting empathy bias share a common goal: to develop a diverse palette of emotions and refine emotional granularity

14 I take Nussbaum's work on morality and literature to be particularly amenable to some of the pragmatist anti-absolutist ideas discussed in Chapter 6. For example, Nussbaum (1985) argues that morality involves "intense scrutiny of particulars" (p. 516), that "to confine ourselves to the universal is a recipe for obtuseness" (p. 526), and that literature offers a particularly effective view of moral life and means of combating such obtuseness because "a responsible action… is a highly context-specific and nuanced and responsive thing whose rightness could not be captured in a description that fell short of artistic" (p. 522).

15 Carroll (2001, 2011) argues that engagement with literary characters is not properly conceptualized as empathetic engagement. I will address his account in my discussion of the imaginative approach.

through the effortful pursuit of diverse experiences. Developing this emotional breadth and emotional granularity combats empathy bias because it enables one to better construct the emotions of others such that one can accurately incorporate their perspectives as evidence in moral inquiry.

With this conception of emotion and the possibility of correcting empathy bias in hand, I turn in Chapter 4 to the specific arguments made by Prinz and Bloom and consider whether the evidence of empathy bias is enough to justify their arguments that we ought to reject a role for empathy in morality. The question of whether we should make such an effort to correct empathy bias, rather than pursue other moral approaches, remains open despite my argument in Chapter 3 that effortful correction of empathy bias is possible. In Chapter 4, I argue that while both Bloom and Prinz are right to be concerned about empathy bias, they are wrong to conclude that empathy should not be a part of our moral lives. The key is that Bloom and Prinz tend to focus on the problems of drawing on empathy with the suffering of others as the primary motivating force behind our moral judgments, while they neglect the potential benefits of empathy as a critical tool to incorporate in the process of critiquing our own moral views and assessing moral problems from other perspectives. In making this argument, I turn to their purposed alternatives: rational compassion for Bloom, and non-empathetic moral emotions for Prinz. Thus, my goal in Chapter 4 is to argue for the idea that despite the prevalence of evidence of the sort described in Chapter 2, and contrary to the views of Bloom and Prinz, empathy provides a unique value to moral inquiry, particularly that of valuable evidence, and that this role is worth preserving because of the value of such evidence in critical moral self-assessment and development.

Bloom argues that empathy should be replaced by rational compassion. I argue that this approach is misguided for two reasons. The first is that Bloom overlooks the potential of

augmenting compassion-driven reason with empathy in order to identify potential egocentric biases in moral reasoning. Bloom's claim is that "while sentiments such as compassion motivate us to care about certain ends—to value others and care about doing good—we should draw on [the] process of impartial reasoning when figuring out how to achieve those ends." Yet an underlying motivation of compassion may lead to different conceptions of what the appropriate ends are in a moral disagreement. Both sides in a moral disagreement may generally "value others and care about doing good," but they may still disagree about the subtleties of the problem and what compassion should lead us to do. This is not a dispute about the appropriate way to apply reason towards an agreed upon compassionate end; it is a dispute about what the appropriate compassionate end is in the situation at hand. It is a dispute about values. My claim is that if we do not make an effort to empathetically understand values that differ from our own but have a legitimate claim to being compassionate, then we are not reasoning with the sort of impartiality that is Bloom's ultimate goal. Empathizing with other perspectives on what the most compassionate solution to a moral problem is allows us to test our own values from a less egocentric perspective and combat tendencies to assume that our proposed solution to a given is the only or best rational, compassionate solution. Such an assumption may be based on our own limited experience, and empathy allows us to widen that experience so as to factor in individual and cultural values that may initially be somewhat foreign to us, but that are nevertheless relevant to the moral problem at hand. Empathy allows us to critique our own view of the appropriate moral ends to seek in specific moral problem situations and to do so based on the perspectives of others involved. As such, empathy can help correct egocentric biases regarding our views on the appropriate solution to a given moral problem.

There is a sense in which Bloom recognizes Hume's point that reason is the slave of the passions. Bloom's point is not that reason operates in absence of emotion, but rather that in terms of moral deliberation, the emotion that ought to motivate reason is a diffuse compassion and not empathetic connection. I am sympathetic to the idea that compassion ought to motivate reasoning in certain cases of moral deliberation. However, my point is that we need some means of testing our preferred compassionate solutions if we are to remain truly impartial. An appeal to "value others and care about doing good" is too vague to be helpful in our actual encounters with moral dilemmas. Which others ought we value? What kinds of goods? Empathy's role in moral inquiry is to help one to explore alternative answers to these questions by considering the emotional perspectives of others involved in the moral debate and how these perspectives relate to one's own values.

This leads to my second critique of Bloom, namely that it is in fact compassionate to empathize with others so as to recognize their status as individuals worthy of nuanced consideration and moral agents whose perspective is relevant to moral inquiry. Compassion motivates us to value equity and the concerns of others, but part of valuing others is valuing their authenticity as moral agents, their status as beings with unique goals and emotions. Rather than imposing one's own moral views on others or treating them as uniform factors in a utilitarian calculus, one ought to remain open to the views, goals, and emotions of others as grounded in their specific experiences. Doing so recognizes the authenticity of other moral agents and can open avenues of communication that are beneficial to moral inquiry. Thus, a motivation to be compassionate ought to motivate us to make an effort to empathize, because this effort is a recognition that an individual is an authentic agent worthy of detailed consideration.

Prinz does not advocate for the sort of utilitarian rationality that Bloom seems to favor. Rather, he argues that empathy is not necessary for moral judgment, behavior, or development and that it is other moral emotions such as guilt, outrage, and compassion, but not empathy, that are the basis of morality.

I present two objections to Prinz's view. First, I critique his objection to "agent empathy" accounts that emphasize the role of empathy with the agent performing a morally salient action in the process of approving or disapproving of that action. I agree with Prinz that we ought to neglect the "constitutional thesis" that empathy constitutes approbation and a lack of empathy constitutes disapprobation, but I distinguish between evaluating an action and evaluating an agent and highlight a role for empathy in evaluating an agent. Empathy allows us to understand that we may approve of an agent's action but disapprove of the agent's motivation, or *vice versa*. This recognition is valuable to moral inquiry in that it helps us critique how we understand the relation between moral actions and moral agents. So, while agent empathy need not be involved in our moral judgment of an action, it is still valuable to moral inquiry because of its ability to help us understand moral agents and thus to understand the relation of our own sentiments of approbation or disapprobation of actions and of agents.

Second, I argue that empathy plays a key role in the development and evaluation of the sort of moral emotions that Prinz favors. My claim is that Prinz focuses only on the potential role of empathy for the suffering of a victim of an immoral act and neglects the role of empathizing with the moral emotions of those who make moral judgments about the actions of oneself and others. Empathetically understanding that the people around us experience complex emotional responses to the actions of others, including our own actions, allows us to appreciate the moral import, rather than merely the causal impact, of those actions. We develop emotion concepts

such as guilt, love, and indignation through an empathetic understanding of others' moral emotions as they relate to us.

In Chapters 5 and 6 I extend my argument for the value of empathy in impartial moral inquiry. My arguments in these chapters draw on insights from the pragmatist method of moral inquiry found in the work of John Dewey (1888/1993, 1916b, 1939, 1945), Jane Addams (1902/2005), and others.[16] In Chapter 5, I examine Smith's empathy-based account in order to draw out some similarities with my own, but also to pose several objections to Smith's account of empathy-based impartiality. In Chapter 6, I address the difficulties facing Smith's account by drawing on insights from pragmatist philosophers regarding the value of fallibilism in inquiry. I argue that we ought to conceptualize impartiality in terms of adherence to fallibilistic method rather than the construction of a static, idealized "impartial spectator."

Roughly, my claim is that to neglect the capacity to empathetically take on the emotions of another person when engaged in moral inquiry is to violate the principle that C.S. Peirce (1898) declares "deserves to be inscribed upon every wall of the city of philosophy: *Do not block the way of inquiry"* (p. 48, emphasis in original). Neglecting our capacity to take on the emotions of others closes down productive avenues of moral inquiry, as empathetically grasping the emotional underpinnings of other people's moral views, especially their views towards one's own character and conduct, is crucial in assessing the merits or shortcomings of those views and thus is crucial in maintaining a healthy, open-minded critical stance towards one's own moral outlook. Inhabiting the emotional perspectives of others is a central aspect of applying to the

16 See, for example, James (1891a), Hilary Putnam (1990, 1992, 2002), and Ruth Anna Putnam (2009). While I focus mainly on Dewey and Addams' moral philosophy, I take the aforementioned philosophers' work to be amenable, and in some cases inspired by, Dewey's and Addams' emphasis on fallibilistic, anti-absolutist, democratic, empirical inquiry in the moral realm. My goal is not to wholeheartedly support all aspects of these philosophers' various accounts, but rather is to draw out what I take to be some central insights that they share, insights that are relevant to my defense of the role of empathy in moral inquiry.

moral realm Dewey's (1910, 1916b) pragmatist conception of intelligence as a fallibilist, empirical, and imaginative exploration of possible solutions.

In Chapter 5, I discuss Smith's sentimentalist account of moral inquiry, which overlaps with my own in that moral inquiry is an empathetic, social, and empirical process. However, I conclude the chapter by highlighting two problems that face Smith's account and thus call into question whether we ought to adopt such an empathy-based approach. These problems arise because of a tension between Smith's commitments to the empirical, social nature of moral inquiry and his appeal to an idealized impartial spectator and general moral principles as means of combating biases. Roughly, the problems are: (1) an empirically constructed impartial spectator constructed based on empathy with the sentiments of others may be constructed according to empathetic biases and thus lead one to a biased conception of ideal impartiality, and (2) disagreements between competing empirically constructed conceptions of what constitutes the perspective of an impartial spectator cannot be settled by appeal to an impartial spectator without begging the question or infinite regress.

In Chapter 6 I address these problems, along with the more general problems of empathy bias discussed in Chapter 2. My goal is to draw on insights from pragmatism to defend the social, empirical nature of Smith's empathy-based account while avoiding the problems that stem from his account of impartiality. I defend an account of empathetic moral inquiry that is influenced by pragmatism in that it shares a foundation with the pragmatist method of inquiry, which is grounded in fallibilism, anti-absolutism, empiricism, and democracy.[17] A commitment to this fallibilistic mindset is linked to a commitment to pursuing diverse experiences via the embedded, communicative, and imaginative approaches discussed in Chapter 3 and thus is a

17 The terms 'democracy', 'anti-absolutism', 'fallibilism', and 'empiricism' are of course understood in a variety of ways. In Chapter 6, I define and use these terms in a manner that is consistent with their use by the pragmatists that I discuss.

commitment to effortfully correcting empathy bias. Furthermore, I argue that, because this pragmatist approach defines impartiality in terms of adherence to fallibilistic method rather than in appeal to an ideal, absolutely impartial spectator, it is able to avoid the problems that face Smith's account.

When the claims of Chapters 1-6 are taken together, we are left with the following picture of the role of empathy in moral inquiry. First, we should understand empathy as involving interconnected cognitive and affective components. Second, although the emotional nature of empathy is rightly taken to be problematic due to its susceptibility to bias, we can overcome this problem by effortful calibration of our empathetic engagement, which involves conscious pursuit of diverse experience and utilizing other-oriented perspective taking so as to cultivate the emotional breadth and granularity required to construct the emotions of a diverse variety of others. One can improve one's empathetic capacity through direct experience, communication, or engagement with works of art that portray diverse, nuanced perspectives. Third, it is worth pursuing the correction of empathy bias, rather than trying to remove empathy from our moral lives, because empathetic engagement is uniquely capable of facilitating impartial moral inquiry precisely because of its ability to allow us to experience, to some degree, the emotions of people who do not necessarily share our own values, and to critically examine our own moral beliefs and conduct from outside of our own perspective. We will block inquiry if we fail to utilize our capacity to experience the emotions of others to some degree; such an experience allows us to assess our own moral perspective from a critical distance, and thus provides us with the mirror necessary to view the propriety or demerits of our own sentiments and conduct.

Chapter 1

Defining Empathy

C. Daniel Batson (2009) identifies eight definitions employed by philosophers, psychologists, and neuroscientists in their work on 'empathy.' Though I will not discuss each of these definitions in detail, over the course of this chapter I will, like Batson, outline some of what I take to be the relevant distinct uses of the term 'empathy' in the literature, with the aim of setting aside uses that are not at issue in this dissertation. I focus especially on distinguishing my account of empathy from emotional contagion, a mere sharing of affect that does not involve any cognitive grasp of the other's perspective. I also want to distinguish the phenomenon I have in mind from theory-theory, a theoretical understanding of others' mental states based on folk psychological laws and not on any simulation of the others' experience. This does not necessarily mean that I take these phenomena to be non-existent; however, it does mean that I take them to be importantly distinct from the capacity that I have in mind in this dissertation, and that I take my conceptualization of empathy to be more relevant to moral inquiry.

I have two overarching goals for this chapter: clarifying exactly what I take empathy to be and highlighting its role in our efforts to understand the emotions of others. I outline and defend an account of empathy based on the work of Amy Coplan (2011a, 2011b), whose "narrow conceptualization" involves both affective and cognitive components. My account also draws on Lisa Barrett's discussion of the role of basic affective valence and context in constructing complex emotions. In defending this account, I discuss and support Alvin Goldman's arguments in favor of the role of simulation in our efforts to understand the minds of others, i.e., Theory of Mind (ToM). The result is an account of empathy as a simulation-based process that necessarily involves both affective and cognitive elements.

Before I proceed, I want to make it clear that my aim here is not to define empathy so as to render it unproblematic in our moral lives; I do not want to conceptualize empathy in a way that leaves it immune to concerns regarding bias. In fact, much of the empirical evidence I will adduce here in defending the existence and significance of the capacity that I have in mind is evidence of the susceptibility for this capacity to be biased and error prone. Nevertheless, I take it that the empirical evidence highlighting the problems we encounter when attempting to understand the minds of others points us away from theory-theory models and offers support for my model of empathy as a simulational process.

The chapter is organized as follows:

In section 1, I draw on Coplan's account to define what I take to be the sort of empathy that should be cultivated and applied in moral inquiry, distinguishing it from a phenomenon that I take to be less relevant: emotional contagion, a mere sharing of affect. I distinguish the sharing of affect from the sharing of emotion and emphasize the importance of the cognitive act of contextualizing a shared affective experience when empathizing with others' emotions.

In section 2, I draw on Goldman's work to defend the empirical claim that simulation-based empathy is something that we do in fact make use of, albeit it not unproblematically, in our efforts to understand the minds of others. My aim is to show that simulation theories of ToM or "hybrid" theories that involve a simulational component, both of which are consistent with my account of empathy, can better explain the relevant evidence than can theory-theory accounts of ToM that claim that we do not simulate others' perspectives in the process of understanding their mental states.

Taken together these two sections should outline the specific account of empathy that will be the focus of this dissertation. My account is distinct from accounts of empathy that

emphasize mere sharing of affect (the phenomenon I call emotional contagion), insofar as my account holds that an understanding of the context in which the other's affective experience is situated is an essential part of simulating the other's emotion. My account is also distinct from accounts that emphasize only cognitive understanding of the other's emotional perspective without any simulational component (e.g., theory-theory accounts), as I argue that there are good empirical and conceptual reasons for thinking of empathy as a simulation-based process.

1.1: Empathy, Affect, and Perspective Taking

1.1.1: Conditions for Empathy

In this section I defend empathy as defined by Coplan (2011a, 2011b), who argues that empathy involves three necessary features: affective matching, other-oriented perspective taking, and self-other differentiation.

Affective matching occurs when the empathizer's affective state qualitatively matches, to some degree, the affective state of the person with whom she is empathizing. When X empathizes with Y, X must feel, to some degree, the way that Y feels.[18] Importantly, I take affective matching, absent any context that comes from considering the target of empathy's perspective, to involve merely matching with what Barrett (2005, 2006, 2011) and others[19] have called "valenced core affect." Barrett (2011) argues that we can understand complex emotions as constructed out of basic valences of the core affective system, which "consists of neurobiological states that can be described as pleasant or unpleasant with some degree of arousal" (p. 363). The idea is that emotions involve certain combinations of pleasantness or

18 One may take oneself to share another's affect but be mistaken, as the target of empathy is in fact not experiencing the feeling that the empathizer thinks that she is sharing. In such a case, though the would-be empathizer believes that she is empathizing, on my account she is not actually empathizing insofar as there is no degree of affective matching. However, my account of empathy is of a phenomenon that occurs in degrees, so one could be said to be empathizing if their simulation matches the target's experience to at least some degree.
19 See, for example, Russell (2003).

unpleasantness with varying levels of arousal, but which emotion is constructed is crucially dependent on the context in which the core affective valence is experienced. On Barrett's view, emotions are multiply realizable and context-dependent. For example, an emotion such as grief may involve high arousal, as when experiencing intense shock at the unexpected loss of a loved one, or low arousal, as when experiencing a more subdued, reflective sadness over a loss. Furthermore, the same affective state of low arousal may be involved in depression or in deep calm, and core feelings of pleasantness will be involved in a wide variety of more complex emotions such as pride, gratitude, and love, with the context in which such pleasantness occurs being a crucial factor in determining which emotion is constructed. These are just some examples, and I will discuss Barrett's account in much more detail in Chapter 3, but for now the important point is that we must consider the *context* in which a person's core affective valence is situated in order to identify an instance of any emotion. Considered in terms of empathy, this means that affective matching must be contextualized in terms of the other's perspective if one is going to empathetically take on the other's emotions rather than merely share the other's core affective valence.[20]

This leads us to the second necessary condition of empathy: other-oriented perspective taking. Other-oriented perspective taking occurs when the empathizer imagines undergoing the experiences of the person with whom she is empathizing from that person's perspective; this is crucially distinct from what Coplan calls "self-oriented perspective taking," in which the empathizer imagines herself undergoing the other person's experiences.

20 Throughout this dissertation I will be careful to distinguish affect from emotion. As will become clear as the dissertation progresses, I take emotions to be complex experiences involving conceptual content, while I take affect to be a mere valanced disposition. Although affect plays a necessary role in the construction of emotion, it only does so insofar as it is contextualized in terms of emotion concepts and the environment in which it occurs.

This means that the empathizer must engage in some level of what Goldman (2006) calls "quarantine" of one's own perspective; there is a sense in which one must not allow one's specialized knowledge or idiosyncratic beliefs and desires to "seep into" (p. 29) the simulation of another's perspective who lacks these states. The extent to which we are able to effectively engage in this sort of quarantine is an empirical question. As we shall see in the following section, the answer suggests both that we do often attempt to engage in other-oriented simulation, rather than understand the minds of others through folk psychological theory alone, but also that we often encounter difficulties in effectively quarantining our own perspective. However, it is important to note at this point that it would be far too strict to require that empathy involve the *complete* taking-on of another's perspective such that the empathizer's perspective becomes indistinguishable from the target of empathy. Rather, like affective matching, other-oriented perspective occurs in degrees. In each case of empathetically simulating the perspective of another, there are relevant aspects of the other's perspective that one must make more of an effort to take on, along with less relevant aspects that one need not make such an effort to consider. In order to empathetically simulate another's emotional response to a particular moral problem, say the responsibility of human beings regarding climate change, I will not need to empathetically simulate your emotional response to some unrelated moral problem such as capital punishment. Furthermore, one need not perfectly simulate even the relevant aspects of the other's perspective; I may take on your perspective regarding climate change to a degree, with the degree corresponding to my ability to quarantine my own views. In Chapters 4 and 6, I argue that this tension between effortful other-oriented perspective taking and varying levels of difficulty in quarantining our own views is helpful to moral inquiry in that it allows us to bring to light and potentially challenge the background assumptions of our own views against the

backdrop of another's experience, but for now it is enough for my definition to emphasize that empathy need only involve *some degree* of other-oriented perspective taking. This perspective taking enables the empathizer to fill out the context in which the other's affective valence is situated. In other words, other-oriented perspective taking enables one to experience a degree of empathy with an emotion rather than mere affective matching.

There is a third necessary feature required for empathy: self-other differentiation, which occurs when one preserves one's sense of self despite engaging in other-oriented perspective taking. Although she imagines experiencing the other person's experiences from that person's perspective, the empathizer remains aware that the other is a separate person with unique thoughts and feelings. The above discussion of degrees of other-oriented perspective taking is again relevant here. If one were to inhabit the perspective of another entirely, it is not clear that one would be able to distinguish oneself from the target of empathy at all. I do not know of any situations in which such a phenomenon occurs, and in any case, it is not what I have in mind when discussing empathy.

These three features are jointly sufficient for empathy. Empathy is the state that occurs if, and only if one experiences some degree of affective matching with another person through engaging in some degree of other-oriented perspective taking while maintaining self-other differentiation.

1.1.2: Emotional Contagion

Other accounts of empathy tend to emphasize one of these conditions over the others, or to distinguish different kinds of empathy based on which of these conditions are met. For example, many accounts distinguish affective empathy, which requires only affective matching, from cognitive empathy, which requires only an understanding of the other's perspective and no

affective sharing. By contrast, my account of empathy is meant to highlight the connection of these two components. Nevertheless, I find the distinction between affective empathy and cognitive empathy to be helpful insofar as it allows us to see what phenomena I do not have in mind when talking about empathy.

First, my account distinguishes empathy, which involves contextualizing a shared affective experience through other-oriented perspective taking, from affective empathy considered as synonymous with emotional contagion, a process that, as Stephen Davies (2011) explains, "involves the transmission from A to B of a given affect such that B's affect is the same as A's but does not take A's state or any other thing as its emotional object" (p. 138). In cases of emotional contagion, one experiences some degree of matching with the core affective valence of another, but until this affective valence is contextualized, one does not actually share the other's emotion on my view, because emotional experience requires consideration of the context in which one experiences a particular affective change.

There is ample empirical evidence documenting the existence of the phenomenon of emotional contagion.[21] It is a real and interesting feature of human psychology that is worth studying, however it is not my focus here. I am concerned here with empathy's role in our moral lives, particularly its role in helping us understand and critique moral emotions throughout the process of moral inquiry, and emotional contagion lacks the consideration of context that is necessary in order to understand an affective experience as part of a moral emotion.

For example, suppose that as I take the subway home from work, I sit next to an individual who has witnessed a hit and run car accident earlier in the day and who is experiencing an emotion of moral outrage as a result. This individual may signal this outrage with certain facial expressions and body posture that I associate with an unpleasant/high arousal

21 See Hatfield et al. (2009) for an overview of this research.

affect, and I may end up catching her affect of high-arousal unpleasantness as she, say, silently fumes with a scowl and crossed arms. However, if I do not understand what this affective experience is related to, in this case the hit and run accident, then I have not gained any sort of morally relevant experience. I do not experience my affective change as relating to a particular morally salient event, and thus do not engage in any sort of introspection about what the appropriate response to such an event would be. Rather, I simply feel unpleasant. As such, this sort of case of emotional contagion does not have the relevance to moral inquiry that I will argue empathy possesses.

In fact, emotional contagion could be problematic for moral inquiry, as catching another's affect and contextualizing it as relating to a moral problem that it in fact is not causally related to may lead us to make rash moral judgments that are not grounded in careful inspection of the relevant contextual factors. For example, Danziger et al. (2011) found that judges were significantly more likely to deny parole in cases that were heard just before lunchtime. Barrett (2018, pp. 74-75) argues that a possible explanation for this is that the judges unwittingly contextualized the unpleasant affective valence involved in hunger as relating to the defendants' cases. This is an example of what Barrett calls "affective realism,"[22] which occurs when, lacking an awareness of the origin of a particular affective valence, one projects one's affective experience onto reality. Unaware that her negative affect is related to hunger, a judge could take the cause of her feelings of unpleasantness to be the defendant's case at hand and judge the defendant according to this negative affective response. In such a case the lack of appropriate contextual understanding of one's affective experience leads to the construction of an inappropriate emotion.

22 See Barrett (2018), pages 75-78.

Considered in terms of emotional contagion and moral judgment, affective realism is problematic in that one can catch another's affect and project it onto moral situations that in fact are not the cause of such an affect. For example, suppose that rather than having an unpleasant affect due to hunger, a judge caught another person's unpleasant affect via emotional contagion just before entering the courtroom and hearing a particular case. The risk is that the judge could then project this unpleasant affective experience onto the defendant's case in a manner that would influence her judgment, even though it in fact was merely the result of encountering someone having a bad day moments earlier.

In sum, given the lack of contextual content involved in emotional contagion and its susceptibility to lead one to make problematic projections based on mere sharing of affect, I will bracket this phenomenon from the account of empathy that I defend in this dissertation. Merely catching another's affective valence without understanding the relevant contextual information surrounding that affect is not sufficient for empathy and does not share empathy's value to moral inquiry.

1.2: Simulation and Theory-Theory

In section 1 of this chapter, I argued that empathy involves taking on the target's perspective. I concluded that in cases of emotional contagion we fall short of empathizing because we fall short in taking on the other's perspective outside of mere affective experience. Nevertheless, I understood both emotional contagion and empathy to be simulation-based processes. In cases of emotional contagion, we do not merely understand that another person is experiencing a certain affective valence; we experience, to some degree, that valence ourselves. In cases of empathy, we do not merely understand that a person has some contextualized

affective perspective; we experience, to some degree, that perspective ourselves. I have been arguing that when we empathize, we *simulate* the experience of the target of empathy.

However, many theorists in the ToM debate argue that in fact we do not employ such simulation-based processes when understanding others, but rather utilize a theoretical understanding of the other's perspective in which behavior is predicted and understood according to folk psychological laws. In this section I will argue, largely based on empirical evidence, that these theory-theory accounts of ToM are mistaken and that ToM is better explained by the sort of simulation-based process that I take empathy to be. In particular, I defend Alvin Goldman's interpretation of relevant empirical evidence as supporting the role of simulation in ToM.

Goldman (2006) provides an extensive defense of simulation theory over theory-theory in the ToM debate, and ultimately settles on what he calls a hybrid account according to which simulation and theoretical understanding both play a role. For my purposes, the important point to emphasize is that, even on such a hybrid account, simulation does play a role. It is this role that I am concerned with in my account of empathy. Examining Goldman's critique of theory-theory will be helpful in understanding why I defend a simulation-based conceptualization of empathy. I will look closely at two of Goldman's arguments. The first involves paired deficits in emotional experience and the attribution of emotion to others. The second involves egocentric projection of one's own mental states to the minds of others. Paired deficits and egocentric projection both provide good evidence that we engage in simulation rather than purely theory-driven processes in our understanding of the minds of others.

Before examining Goldman's arguments in detail, it is important to clarify the distinction between simulation theories and theory-theories. Though there are of course a variety of subtle differences amongst the views of theory-theorists, for my purposes here it is enough to highlight

the view that I take all of these accounts to share, namely that in order to understand the minds of others, we employ lay psychological theories about human behavior that utilize folk psychological laws. So, for example, on a theory-theory account, my belief that you desire X is not based on any sort of simulation of your desire, but rather is based on my use of a psychological theory according to which people in your sort of situation with your knowledge and beliefs tend to desire X. By contrast, simulation theories stress that our understanding of the minds of others necessarily involves some level of simulation of others' experiences. According to simulation theory, my belief that you desire X will involve some level of simulation of your desire; I experience your perspective rather than merely theorize about it.

We are now in a position to evaluate evidence as supporting either a theory-theory account of ToM or an account that incorporates simulation. In doing so, my goal is to argue that an account that incorporates simulation, whether hybrid or strictly simulation-based, provides a better explanation of the evidence at hand, suggesting that we in fact engage in simulation-based empathy rather than understand other minds by mere theorizing in the manner described by theory-theory.

1.2.1: Paired Deficits as Evidence of Simulation

A key portion of evidence in this debate, as noted by Goldman (pp. 115-125), is the existence of paired deficits in one's ability to experience a particular emotion and one's ability to attribute that emotion to others. Goldman argues that evidence of paired deficits suggests that an individual's ability to experience an emotion plays a central role in understanding other individuals' experiences of that emotion. This sort of paired deficit is precisely what simulation theory should predict. If understanding your experience of fear necessarily involves me simulating your fear, then my ability to understand your fear will suffer if I have difficulty

experiencing fear; my deficiency in experiencing fear will constrain my ability to simulate fear, and thus, on the simulation theory model, my ability to understand your fear will suffer. This simulation-based explanation is an intuitive explanation of paired deficits.

On the other hand, it is not as clear that theory-theory can provide a good explanation for paired deficits. As Goldman notes (pp. 119-124), in order to explain a paired deficit, theory-theory would have to postulate that a deficit in experiencing some emotion E is accompanied by some deficit in obtaining facts or in engaging in theoretical reasoning regarding E. While this sort of explanation is not impossible, it faces problems in accounting for the *specificity* of paired deficits; that is, it must explain why an individual who experiences a deficit in some particular emotional experience is deficient only in attributing that emotion to others and is not deficient when attributing other emotions. The simulation theory explanation for this aspect of paired deficits is straightforward: if an individual is only impaired regarding the experience of emotion E, then this will not affect her ability to attribute other emotions, as she is capable of experiencing and thus simulating those emotions. On the other hand, as I see it, the theory-theory account needs to defend at least one of two claims[23] in order to explain the empirical evidence regarding the specificity of paired deficits: (1) paired deficits involve deficits in the experience of some emotion E paired with methodological error in theoretical reasoning about E and only about E. (2) Paired deficits involve deficits in the experience of some emotion E paired with problems regarding evidence recognition about E and only about E. Both claims are problematic.

23 As Goldman points out, there is a third possible route that a theory-theorist might take. One could argue that the part of the brain that is damaged in paired deficit patients is responsible both for the experience of emotion E and for the attribution of emotion E, but that these two functions are not related; they merely happen to be carried out in the same area of the brain. While we cannot rule this out in theory, it is highly improbable that such a chance correlation would occur for a variety of different emotions. The theory theorist pursuing this line of argument would have to claim that this coincidental colocalization of emotion attribution exists in a variety of locations throughout the brain and for multiple emotions (e.g., anger, fear, disgust). The burden of proof here lies squarely on the theory-theorist to explain why such coincidental colocalization would be so prevalent, and I know of no explanation that adequately does so.

In order to defend (1) the theory-theorist must give up on a domain-general account of theoretical reasoning and provide an explanation for how our *method* of reasoning varies according to the sort of emotion we are trying to predict. If we reason about different emotions using the same underlying method, a deficit in theoretical reasoning regarding some emotion E should correspond to deficits in attributing a variety of emotions, given that the method for emotion attribution is the same across those emotions. Again, this is not what studies on paired deficits find. For example, as Goldman discusses (p. 116), studies on NM, a patient who suffered damage in the bilateral amygdala and exhibited abnormally low experiences of fear, found that NM performed significantly worse than controls regarding fear attribution, but did not perform worse than controls in attributing happiness, sadness, surprise, disgust, or anger.[24]

Faced with this sort of result, the defender of (1) must explain why NM's deficit in the experience of fear causes him to reason differently regarding fear and only regarding fear when attributing emotions to others. On the theory-theory model, a theorizer about another's mental states takes as inputs various claims about that other's knowledge, beliefs, desires, etc. and draws conclusions based on these inputs and their relation according to folk psychological laws. For example, one's theory may involve a claim such as "if X screams in the context of a potentially harmful situation and believes that she is in danger, then X is experiencing fear." So, on the theory-theory account, given knowledge about when people normally experience fear, we reason that if such conditions obtain for some particular individual, then that individual will feel fear. The defender of (1) must explain why this sort of basic conditional inference breaks down only in cases relating to one particular emotion, and it is not clear why the *structure* of such an inference would be affected by a lack of experience. Indeed, a study by Adolphs et al. (1999) on SM, another patient with bilateral amygdala damage and deficits in the capacity to experience

24 This result is from Sprengelmeyer et al. (1999). See Adolphs et al. (1999) and Adolphs (2002) for similar results.

fear, found that she experienced no deficit in this sort of conditional reasoning. As Adolphs et al. note,

> "(a)t a purely intellectual level she knows what fear is supposed to be, *what should cause it, and even what one may do in situations of fear*, but little or none of that intellectual baggage, so to speak, is of any use to her in the real world." (1999, p. 66, emphasis mine)[25]

So, the theory-theorist lacks support for (1). She encounters difficulty in explaining why a method of theoretical reasoning is deficient regarding some emotions but not others in individuals with paired deficits. But this leaves open claim (2). Perhaps experiential deficits do not impair one's capacity for general theoretical reasoning about some emotion, one's ability to make the sort of law-based inferences required according to theory-theory; rather, experiential deficits impair one's ability to obtain the relevant evidence necessary to furnish one's theory with the sorts of facts that would allow one to accurately attribute that particular emotions to others.

The defender of (2) must show that a deficit in experiencing some emotion E corresponds with an impaired ability to recognize signals of E and only E. However, it is not clear why a deficit in experiencing, say, anger, would correspond to a deficit in one's ability to observe signals relevant to anger, but would not impact one's ability to observe evidence about other emotions. The basic perceptual abilities required to observe the sort of evidence relevant to emotions, things like posture, facial expression, surrounding context, and vocalization, are not impaired in individuals who experience paired deficits. While different expressions, postures, etc. are involved in the recognition of different emotions, the basic types of signals involved in

25 See Calder et al. (2000) for a similar finding regarding a patient with a paired deficit involving disgust, and Lawrence et al. (2002) for a similar finding regarding paired deficits in anger.

emotion attribution are consistent, and paired deficit patients do not lack the ability to perceive these types of information, as evidenced by their normal performance in the attribution of other emotions. Furthermore, as Goldman notes (pp. 120-121), studies of paired deficit patients routinely include standard tests that measure one's ability to visually process faces, and the patients displayed no deficits in tasks such as categorizing different views of faces as belonging to the same face, or recognizing high-level properties of faces such as age, gender, and identity. This suggests that paired deficit patients are capable of recognizing facts about the facial configuration of others experiencing E, but they simply do not recognize these facts as relevant to emotion E. Simulation theory can explain this by pointing out that more than mere recognition of such facts, namely simulation, is required for emotion attribution. Theory-theory, however, appears at a loss. On a theory-theory account, given that one's method of theoretical reasoning about E is not deficient, and that one is capable of obtaining facts pertaining to theories about E, one should be able to appropriately attribute E, yet paired deficit patients cannot do so.

It is not the case that paired deficits involve an inability to experience an emotion paired with a global deficiency in emotion recognition, and yet on the theory-theory account our general method for creating and utilizing theories about different emotions is the same regardless of which emotion we are trying to recognize: we make observations and make inferences based on folk psychological laws. Given that these capacities to make emotion-relevant observations and engage in law-based inference are not globally deficient in individuals with paired deficits, the theory-theorist must explain what it is about a particular experiential deficit that leaves these capacities deficient in only the specific cases relating to the emotion in question. The simulation theorist seems to have a good explanation for this specificity: because our method for emotion recognition involves simulation, a deficit in the ability to simulate a specific emotion will leave

our method deficient only in cases in which the simulation of that emotion is required. This sort of explanation is not available to theory-theorists. A theory-theorist cannot claim that an inability to experience some emotion will only affect our ability to observe evidence and make law-based inferences about that specific emotion, because the basic application of our capacities to observe evidence and make law-based inferences does not vary across different emotions, thus we should see a more global deficit if these capacities are impaired at all.

1.2.2: Egocentric Projection as Evidence of Simulation

As noted earlier in this chapter, there is significant evidence that when attributing mental states to others, we often experience difficulty in quarantining our own mental states and thus wrongly project our own knowledge, beliefs, desires, etc. onto those whom we are trying to understand. In this section I will describe some of this evidence and argue that it provides support for accounts of ToM that incorporate simulation rather than theory-theory accounts.

I will divide my discussion of egocentric bias into two categories: projection of knowledge and projection of feelings. I will highlight the general pattern of egocentric projection that holds across these categories and argue that egocentric biases in projection in general are best explained by the role of simulation in ToM. I conclude the chapter by briefly considering how these results are relevant to my broader project of correcting empathy bias and applying empathy in moral inquiry.

Egocentric bias in knowledge projection occurs when one fails to quarantine one's knowledge, as opposed to beliefs, desires, etc. when attributing knowledge to other people. For example, Camerer et al. (1989) conducted a study in which people who were well-informed regarding corporate earnings were asked to predict how less informed people would forecast the same corporate earnings. The study was designed to reward the more informed people if they

were able to quarantine their own knowledge and accurately predict the earnings predictions of the less informed participants. Importantly, the well-informed participants were made aware that the individuals whose behavior they were tasked with accurately predicting were not well-informed regarding corporate earnings forecasts, thus the well-informed participants should have made an effort to quarantine their own knowledge. However, the well-informed participants were unable to completely quarantine their own knowledge and in fact predicted that the less-informed participants were much more knowledgeable than they actually were.

Goldman describes another study conducted by Newton (1990) in which participants were assigned roles as either "tappers" or "listeners." The tappers were asked to tap the rhythm from a well-known song while the listeners, who lacked knowledge of what song the tapper had chose to tap, listened. Tappers were then asked to predict the frequency with which listeners would be able to guess the song in question. Newton's findings suggest that the tappers were unable to quarantine their privileged knowledge of the songs: while tappers predicted a listener success rate of 50 percent, the actual listener success rate was less than 3 percent.

These are just two particularly interesting cases of egocentric bias in knowledge projection, but there is a significant amount of empirical evidence supporting the phenomenon.[26] The important point for our purposes here is to highlight the pattern that these cases show. When attempting to attribute knowledge to others, we tend to, at least partially, attribute our own privileged knowledge to other individuals, even when we are told that these individuals lack this knowledge. This is especially clear in Camerer et al.'s study. The well-informed participants were told that the less informed participants did not share the relevant knowledge of corporate earnings, yet they still were unable to prevent their own possession of the relevant knowledge from seeping into their predictions of how much the less informed participants knew. If our

26 See Nickerson, Butler, and Carlin (2009) for an overview of this evidence.

account of ToM involves simulation, this is easily explained. In attributing knowledge to the less informed participants, the well-informed participants simulate the less informed perspective, but they encounter difficulty in entirely quarantining their own privileged knowledge and thus wrongly conclude that this knowledge is a feature of the perspective that they are simulating. On the other hand, it is not clear how a theory-theory account can make sense of this egocentric knowledge projection, because the well-informed participants were explicitly given the appropriate information regarding the less informed participants' levels of knowledge. If the well-informed participants were utilizing psychological theories in predicting the earnings forecasts of the less informed participants, then they presumably should have incorporated their awareness of the less informed participants' lack of knowledge as relevant data for theorizing and excluded their own knowledge as irrelevant. On a theory-theory account, the theorizer should only apply the relevant psychological law to the person whom the theorizer is attempting to understand, and in this case that person's knowledge is explicitly different than the knowledge possessed by the theorizer. We might operate with a law such as "The accuracy of person X's prediction will correspond to the degree of knowledge Y that X possesses." Given that the well-informed participants were given Y, it is not clear why they would attribute predictions that did not accurately correspond to the less informed participants' degrees of knowledge if knowledge attribution involved only theoretical inference based on this sort of law.

Importantly, we also see a pattern of egocentric projection in the attribution of feelings. For example, Van Boven and Loewenstein (2003) conducted a study in which some participants first engaged in vigorous exercise then were asked to imagine a situation in which hikers are lost in the woods with no food or water. Other participants did not exercise but were asked the same questions. The study found that participants who engaged in exercise prior to imagining the

hikers were significantly more likely than other participants to claim that the hikers would be bothered by thirst more than by hunger. The proposed explanation for this is that, as a result of the exertion of exercise, participants who exercised were experiencing more thirst than were participants who did not exercise; they then allowed their current experience of thirst to seep into their simulation of how the hikers would feel in an entirely different context. Again, we see here a failure to completely quarantine one's own experiences while attempting to simulate the experiences of another person.

Theory-theories are especially ill-equipped to explain egocentric projection regarding feelings, as it is not clear why experiencing a feeling should influence the cognitive process of theorizing about the strength of that feeling in others. On a theory-theory model of ToM, when I theorize about how you will be more bothered by thirst than by hunger, I do so based on some folk psychological law according to which individuals with your particular desires, beliefs, experiences etc. will be bothered by thirst more than by hunger. An argument is needed to show why my own experience of thirst while engaging in this sort of theoretical inference should influence my application of the psychological law according to which I draw my conclusion about *your* felt experience. Such a law would be meant to generalize about the feelings experienced by individuals in your particular situation, with your particular beliefs, desires, etc.; it would not be meant to generalize about individuals in my different situation, with my different beliefs, desires, etc. Indeed, even if we did happen to share similar background conditions, my personal feelings at the time of theorizing are irrelevant to the act of making a law-based inference regarding your situation.

Of course, I am speaking of ideal application of this law-based theory, and perhaps a theory-theorist will argue that while one's feelings should not influence how one applies a

psychological law to the understanding of another person's feelings, that does not entail that one's feelings cannot influence such theorizing. In other words, a theory-theorist may argue that our feelings can lead to egocentric projection in the form of faulty theorizing, but this does not mean that we are not theorizing, it only means that we are not theorizing properly. We may grant the theory-theorist this but point out that examining how such faulty theorizing gets off the ground points to a role of simulation in theory formation. Suppose we grant the theory-theorists that egocentric projection of feelings is a result of faulty theorizing and is explained in terms of an over-generalization in our understanding of others' mental states. For example, in factoring one's own felt experience of thirst in condition X into theorizing about another person's felt experience of thirst in some different condition Y, one is making the generalization that thirst will be felt in a similar way across these different conditions. Given that, *ex hypothesi*, this inference is motivated by the strength of the theorizer's own felt experience and not by strong inductive evidence, this is a faulty egocentric inference, but the theory-theorist will argue that it is still a theoretical inference that need not involve simulation. However, at this point we must ask why a theorizer would make the inferential leap from the fact that she is experiencing strong thirst in condition X to the idea that another person is experiencing strong thirst in condition Y. Again, given the lack of inductive evidence, it seems that the best explanation for why this inference gets off the ground is that the theorizer engages in some level of simulation of what she takes to be the thirst of the other person, and that she takes this simulation to justify the inference that the other is experiencing analogous strong thirst only because the theorizer's simulation has been contaminated by her own strong thirst. So, while we may be able to provide an explanation of egocentric projection of feelings in terms of faulty theorizing, the faulty inference at the heart

of the theorizing is ultimately grounded in a simulation that is not appropriately quarantined, meaning that simulation still plays an important role in how we understand the feelings of others.

My goal in discussing egocentric projections in this chapter has been to argue that, like paired deficits, egocentric projections offer empirical evidence of the simulational character of our efforts to understand others. However, it is clear that this evidence is also evidence of the deficiencies of our simulations. These deficiencies are especially problematic in the moral realm. Our understanding of the moral experience of others involves understanding their knowledge and feelings, thus egocentric projection of our own knowledge and feelings has the potential to limit the scope of our understanding of diverse moral experiences that could challenge our own views and benefit moral inquiry. Yet, I highlight this problem in the context of defending simulation-based explanations of egocentric bias with the aim of showing that the solution to the problem of biased understandings of others will not be found in removing simulation from the process of mentalizing. The empirical evidence suggests that understanding the minds of others is at some level a fundamentally simulation-based process, thus we cannot simply will ourselves to disengage simulation when trying to understand others, but rather must seek to improve the way that we simulate others' perspectives. As such I take my defense of the role of simulation in ToM both to support my account of empathy as contextualized emotional simulation and to motivate the broader project of this dissertation, namely to present a method by which we can remedy problematic biases in the way that we simulate the perspectives of others, not by avoiding simulating those perspectives, but rather by becoming better simulators in order to realize the value of empathy in moral inquiry.

While I have briefly touched on some of the implications of empathy for moral inquiry throughout this chapter, my main aim has been to articulate and defend a definition of empathy

so as to be in a better position to examine its moral import in the remaining chapters. I have argued for the following claims:

(1) Empathy is a simulational process in which one takes on the emotional perspective of another via effortful other-oriented contextualization of some degree of affective matching. This process necessarily involves interrelated cognitive and affective components; it is not a mere sharing of affect but rather a sharing of affect that takes into account the context in which that affect is experienced by the target of empathy.

(2) There is strong evidence that this sort of simulational process, rather than mere theorizing, plays an essential role in the way in which we actually do attempt to understand the minds of others. Evidence of paired deficits in emotional experience and emotion attribution suggests that simulating a particular emotion is central to recognizing that emotion in others. Evidence of egocentric projection suggests that simulation plays a key role in our attribution of knowledge and feelings to others. Regardless of whether we ultimately favor a pure simulation theory or a hybrid theory in ToM, simulation has a role to play, and it is the role of this empathetic simulation in moral inquiry that will be my focus.

I take it that (1) provides a thorough definition of what I have in mind in discussing empathy for the remainder of the dissertation, while (2) supports this sort of definition as an empirically viable account of how we understand the minds of others, particularly their emotions. This chapter has been concerned with answering two descriptive questions. The first is the question of what empathy is, and the second is the question of whether we do in fact utilize empathy when trying to understand others. I have answered the first question with my account in 1.1, though this account will be developed in more detail in chapter 3, in which I defend a

constructionist theory of emotions and explain empathy in terms of that theory. I have provided support for an affirmative answer to the second question in 1.2.

While I will move on to questions of how we should cultivate and utilize empathy in moral inquiry over the course of the Chapters 3-6, I first need to address a question that has been hinted at here in my discussion of egocentric projection. It is the question of whether empathy is in fact problematically, perhaps even irredeemably, biased.

Chapter 2

Evidence of Empathy Bias

In the previous chapter, my goal was to establish a theory of empathy and adduce empirical evidence suggesting that empathetic simulation plays a central role in our understanding of the perspectives of others. In this chapter my goal is to highlight the problems to which empathizing can give rise, particularly in the moral realm. In order to do so, I will discuss a variety of empirical studies that demonstrate empathy's propensity for bias. This sort of evidence is often a starting point for critics of empathy, and thus it is important to understand it prior to arguing that empathetic bias can be corrected (Chapter 3) and arguing in favor of a role for empathy in moral inquiry (Chapters 4-6).

My goal is not to refute the empirical evidence of empathy's susceptibility to problematic biases, but rather is to motivate the need for the argumentative work that will follow in subsequent chapters. Critics such as Bloom and Prinz are right to be concerned about empathy's susceptibility to bias. The evidence for empathy bias is clear and convincing and it ought to be addressed before critiquing the views of those who reject a role for empathy in our moral lives. So, before turning to my defense of the value of empathy in moral inquiry, we should now examine the empirical case against empathy.

The structure of the chapter is as follows. In 2.1 I distinguish between different varieties of empathy bias and clarify what sort of evidence is relevant to the debate with which I am engaged in this dissertation. The subsequent sections address each of these varieties of empathy bias in turn: 2.2 focuses on intergroup empathy bias; section 2.3 focuses on a bias of scope, or empathy's innumeracy; and section 2.4 focuses on empathy's bias towards those who are more proximate to the empathizer.

2.1: The Varieties of Empathy Bias

There has been a surge of empirical work on empathy in recent decades, and while much of it has touted the benefits of empathy,[27] a significant amount has focused on drawing out the problematic biases in our empathic responses to others. However, as discussed in the previous chapter, it is important to note from the outset that 'empathy' is used in a variety of ways in research on empathy, and this includes research on empathy bias. Some studies focus on a bias in empathy defined as something analogous to care for others, often referred to as "empathic concern," and do not focus on empathy as perspective-taking. Other studies investigate bias in perspective-taking, and still others focus on the relationship between biases in perspective-taking and biases in concern for others. I am concerned here with the latter two sorts of studies.

In order to clarify the distinction between various uses of 'empathy' in research on bias, consider the following example. Simas et al. (2020) found that empathic concern is biased towards members of one's own political party and that this biased empathy may exacerbate partisan division. They present this finding as counter evidence to the view that empathy can be a means of lessening partisan extremism. It is important to note, however, that their results are only considered in terms of "empathic concern," a measurement in the Interpersonal Reactivity Index (IRI) based on participant responses to claims such as "I often have tender concerned feelings for people less fortunate than me." This is distinct from the IRI's measurements for "perspective taking," which is based on participant responses to questions such as "I sometimes find it difficult to see things from the 'other guy's' point of view." Results regarding "empathic concern" need not entail similar results in terms of "perspective taking," the sort of simulation-based phenomenon that I have in mind. While it may be the case that those with higher ratings of empathic concern tend to show more partisan bias, this does not necessarily mean that those with

27 For surveys of this sort of evidence, see de Waal (2009), Rifkin (2009), and Batson (2011).

higher ratings of perspective taking also will show higher ratings of partisan bias. I bring up this case only to highlight that we must be very careful to discern how empathy is measured and discussed in studies purporting to show empathy bias. My goal is to discuss only studies that relate to the perspective-taking character of empathy. I will highlight only work that demonstrates biases in our ability to take on the perspectives of others, including simulation of affect.

With this in mind, I take it that there are three distinct kinds of empathy bias that pose significant obstacles for the effective utilization of empathy in the moral realm. The first is intergroup empathy bias, a tendency to favor empathizing with members of perceived social ingroups over perceived outgroups; this includes, for example, racial, ethnic, national, political, and religious groups. Intergroup empathy bias facilitates prejudice and moral complacency in that it leaves us less likely to understand the perspectives of those who do not share our particular sociocultural background, and to do so based on morally irrelevant factors of group membership. This bias can leave us cold to the specific moral concerns of social groups of which we are not members, even if upon consideration we may take those concerns to be analogous to our own, or to be of greater significance than the concerns of our own groups; the key is that intergroup empathy bias can dampen our capacity to engage in such consideration. In addition, our tendency to favor empathizing with those who come from similar groups may lead us into a reinforcing feedback loop in which we empathize with individuals who tend to share our own views and thus are not challenged to assess our own moral views and test other potential approaches. In other words, we are left with a form of confirmation bias in moral inquiry.

The second form of empathetic bias is what I will call a bias of scope, a bias towards empathizing with specific individuals at the expense of ignoring larger groups. As Bloom (2016)

puts the point, "[empathy's] spotlight nature renders it innumerate and myopic: It doesn't resonate properly to the effects of our actions on groups of people, and it is insensitive to statistical data and estimated costs and benefits" (p. 31). Empathy's bias of scope renders us more likely to focus on specific individuals, even in instances in which this focus leaves us unable to accurately assess the scope of the moral problem at hand and the potential for more broadly effective solutions. To use an example discussed by Bloom, following the mass shooting at Sandy Hook Elementary School in 2012, the town of Newton, Connecticut was inundated with so many donations of toys, gifts, and money that town officials had to ask people to stop donating. Bloom notes that more school children were killed in shootings in Chicago in 2012 than were killed in the Newton massacre. Furthermore, thousands of children die each year from violence in countries outside of the U.S.[28] The relevantly affluent community of Newton received millions of dollars in donations after Sandy Hook, yet donations to help children and families struggling with poverty in Chicago or in war-torn countries did not increase following the Sandy Hook shooting. Rather, the overwhelming empathetic connection that so many Americans had with the individual parents and family members of victims that they saw grieving during coverage of the aftermath of the shooting triggered a particularly strong moral response directed at those particular individuals. This is not to say that these individuals did not deserve moral consideration; of course they did. Rather, it is to say that our tendency to empathetically engage with individuals and smaller groups can blind us to a broader view of the very issue with which we are morally concerned, in this case violence against children, and thus bias our response to the problem in a manner such that it is less broadly effective.

The third empathetic bias is a bias of proximity and exposure. This is similar, but importantly distinct from the bias of scope. While a bias of scope can often be triggered by

28 See the World Health Organization's Global Status Report on Preventing Violence Against Children (2020).

proximity or exposure, as in the case of the large amount of media coverage of the Sandy Hook shooting, a bias of proximity or exposure is not necessarily a bias toward favoring individuals or smaller groups over larger groups. Rather, it is a bias towards favoring the individuals or groups to whom we are exposed, either via arbitrary geographic proximity or via media coverage, over other individuals or groups that are more or equally entitled to our moral concern. The bias causes us to ignore distant but morally relevant individuals or groups merely because of our empathetic focus on the individuals or groups to which we happen to have more proximity and exposure. Because empathy requires direct interaction with or exposure to the individuals with whom we empathize, we will be more likely to empathize with those whom we happen to have more media exposure to or to whom we are more geographically proximate. But geographic proximity or levels of media exposure should not be a relevant factor in most moral judgments, for example judgments such as considering which charities address the most pressing needs, or determining which people require the most humanitarian relief from the negative impacts of climate change or the Covid-19 pandemic.

In sum, empathy bias can lead us to favor those who are similar to us, to favor individual concerns over group concerns, and to favor those to whom we happen to have more exposure or proximity. There is empirical evidence for each of these biases. I will discuss each in turn.

2.2: Intergroup Bias

Of the three types of empathy bias discussed above, intergroup empathy bias is perhaps the most well-documented. Intergroup empathy bias has been demonstrated in terms of racial, partisan, national, religious, and socioeconomic group membership. Studies have even found

intergroup empathy bias amongst sports fans.[29] Studies have demonstrated intergroup empathy bias amongst arbitrarily created groups.[30] Thus, when we consider intergroup empathy bias as a whole, the takeaway is that intergroup empathy bias is not a feature that is unique to one kind of social group or one kind of individual. Rather, the diversity of groups that demonstrate intergroup empathy bias suggests that this bias is a feature of the very nature of empathy and intergroup dynamics. In order to see this, it will be helpful to discuss two different kinds of intergroup empathy bias: racially-based empathy bias, and partisan-based empathy bias. My goal in discussing the following studies is both to highlight the fact that intergroup empathy bias is involved in a variety of intergroup interactions involving different kinds of groups and to draw out some of the problematic implications of this bias for the role of empathy in morality.

2.2.1: Racially-Based Empathy Bias

Given the global history of racism, it is unfortunately perhaps not surprising to find evidence of intergroup empathy bias regarding racial groups. When one reflects on racial violence and discrimination, it is hard to imagine that the perpetrators of such immoral acts were truly able to empathize with their victims, to inhabit their perspectives. Indeed, we often take such inhumane behavior to be evidence of a lack of empathy.[31] However, it is important to recognize that racially-based empathy bias is not limited to the extremities of slavery, explicit discrimination, and widespread racially-motivated violence; rather, this bias is often implicit and serves as an undercurrent for less overt discrimination that is nevertheless morally unjust. In the United States, we need look no further than the impact of mass incarceration, police violence, and the unaddressed long-term effects of housing discrimination on African Americans to

29 See Cikara et al. (2011c).
30 See, for example, Masten et al. (2010).
31 See Jeske (2018) for case studies that explore this idea.

recognize this implicit racial bias.[32] I bring up this sort of implicit bias because, while some may argue that it is too much of a leap to argue from controlled laboratory experiments to the prevalence of behavior in the real world, I take it that the implicit racial discrimination that we see today lends credence to findings of implicit intergroup empathy bias amongst racial groups in the laboratory. With this in mind, we can now consider a few examples of studies that demonstrate intergroup empathy bias amongst racial groups.

Some researchers have looked to neuroscience to demonstrate this bias. For example, Xu et al. (2009) found that participants showed greater activity in the anterior cingulate cortex (ACC), an area of the brain that is activated in the experience of pain and in the witnessing of pain,[33] when watching members of racial ingroups experience pain than when watching members of racial outgroups experience pain. In the study, researchers used functional magnetic resonance imaging (fMRI) to track ACC activity in Caucasian and Chinese participants while they watched video clips of Caucasian or Chinese faces receiving painful stimulation in which the face was penetrated by a needle, or non-painful stimulation in which the face was touched by a cotton Q-tip. They found that ACC activity was higher in participants when viewing painful stimulation in members of their own race. This result held for both Caucasian and Chinese participants.

Recall that I have defined empathy as involving both affective matching, considered in terms of some degree of matching with a valance of pain or pleasure and arousal, as well as contextual other-oriented perspective taking. Xu et al.'s study seems only to measure affective matching, namely in terms of matching with the target's valence of pain.[34] However, affective matching is a key component of empathizing. While this affective matching must be contextualized in order to count as empathy on my view, it is nonetheless significant that racial

32 See, for example, Alexander (2012) and Coates (2014).
33 See Singer et al. (2004); Botvinick et al. (2005); Jackson et al. (2005); and Saarela et al. (2007).
34 See Mathur et al. (2010) for a similar finding.

group membership can moderate the degree to which one is able to affectively match with another, as this degree is a key factor in the degree of empathy experienced. A diminished affective match based on race can lead to diminished empathy that is ultimately grounded in the morally irrelevant factor of the target of empathy's race.

Nevertheless, studies that measure more contextual and realistic perspective taking scenarios will do more to show empathy bias on my view. In this vein, Johnson et al. (2002) found that white participants were less able to empathize with Black defendants in criminal cases than with white defendants. Furthermore, white participants were more likely to recommend harsher sentences for Black defendants than for white defendants who committed the same crime. Participants in the study were provided with a scenario in which a criminal is on trial for grand larceny but expresses remorse and contextual explanation in a personal statement, which the participants were asked to read. After reading the statement, the participants were given prompts[35] designed either to induce high-empathy (e.g., "try to imagine how the defendant felt" and "try to put yourself in his position"), or induce low-empathy (e.g., "try to be as objective as you can"), while other participants were not given any prompts to empathize (the "no-empathy" condition). Participants then answered a survey in which they rated their empathetic response to the defendant. Johnson et al. found that empathy ratings for the white defendant were higher than empathy ratings for the Black defendant across the high, low, and no empathy conditions. Crucially, the punishments recommended for Black defendants were also more severe than those recommended for white defendants across these conditions. While inducing empathy did induce white participants to make less severe punishment recommendations for Black defendants in comparison to the no empathy condition, these recommendations remained more severe than those recommended for white defendants in the same empathy condition. Thus, the white

35 These were drawn from Batson, et al. (1997).

participants in the high empathy condition were still more likely to recommend harsher sentences for Black defendants than for white defendants, suggesting that they were more easily induced to be empathetic for a white defendant than for a Black defendant.

In contrast to Xu et al.'s study, which is more removed from real-word experience, Johnson et al.'s study demonstrates empathy's susceptibility to bias in a context that is both common in the real world and has a monumental impact on those who are victims of bias; if white jurors experience intergroup empathy bias, then this can result in harsher sentences based purely on race. As noted above, the current state of the United States prison system, in which Black inmates substantially outnumber white inmates, is consistent with this analysis.[36]

Johnson et al.'s study speaks to the possibility that racially-based intergroup empathy bias can lead to racially-based injustice, but justice is not the only moral principle at risk. There is also evidence that racially-based intergroup empathy bias is related to racially-based differences in altruistic behavior. For example, in a 2007 study, Cuddy et al. asked Black and white participants to infer the emotions of individual Hurricane Katrina victims of different races and report their intentions to help these victims. The study found that participants attributed fewer "secondary, 'uniquely human' emotions (e.g., anguish, mourning, remorse)" (p. 107) to outgroup victims than to ingroup victims, and that participants were more likely to help an ingroup victim than an outgroup victim. Given the role of empathy in inferring emotions in others, this result suggests that a bias against empathizing with members of other races can lead to a bias against helping members of other races.

36 Of course, as Alexander (2012) argues, there are a variety of societal, psychological, and political factors at work in racial discrimination within the U.S. justice system. I do not mean to suggest that we can attribute sentencing disparities entirely to empathy bias, only that such disparities are consistent with findings of intergroup empathy bias in sentencing contexts.

Importantly, we see bias in this study occur in terms of an ability to attribute secondary emotions, emotions which Barrett would call more "fine-grained." Emotions such as anguish, morning, and remorse are more complex than mere sadness or unpleasant affect; empathizing with these more complex emotions requires consideration of the context in which they occur from the perspective of the person that is experiencing them. As such, a bias against inferring these more complex emotions in outgroup members suggests a bias against considering the particular context in which those outgroup members experience emotion. To put the point in terms of my conceptualization of empathy, Cuddy et al.'s and similar studies[37] suggests intergroup empathy bias not only involves a bias in affective matching, as shown in Xu et al.'s study, but also involves a bias in other-oriented perspective taking. We see in this case the dangers that empathy bias presents in the form of blinding us to contextual considerations that can and should influence our moral decision to help.

In sum, racially-based intergroup empathy bias can occur across a range of affective and contextual perspective-taking activities, and the results are morally problematic. This bias can lead to injustice and a failure to help those in need.

2.2.2: Partisan Empathy Bias

Just as studies showing racially-based intergroup empathy bias align with historical and contemporary evidence of racial discrimination, studies showing politically-based empathy bias align with our experience of an increasingly politically polarized world. This political partisanship has been acutely felt in the United States, and can be seen not only in voting behavior and political media coverage, but also in increases in political violence. It is unsurprising that increasing partisanship in the United States and other countries has correlated

37 See, for example, Leyens et al. (2000).

with the rise of cable news outlets and of social media platforms such as Facebook and Twitter.[38] These platforms allow users to ignore news and opinions that they consider unfavorable to their own views, and to engage only with news and opinions that confirm their own political and social values. Considered in terms of intergroup empathy bias, social media platforms are dangerously equipped to facilitate the confirmation bias feedback loop discussed above. While social media users have the option of engaging with a diverse set of opinions and views, intergroup empathy bias causes many users to empathize only with the limited perspectives of members of their own political group, thus seeking out and confirming only the views held by that group without leaving open the possibility of debate or revision based on the perspectives of members of perceived outgroups. My aim in drawing attention to this political situation is to suggest that studies that purport to demonstrate partisan-based intergroup empathy bias in a controlled setting are bolstered by the clear hyper-partisanship that we find in the real world, particularly as this hyper-partisanship appears to be related to social media platforms that are uniquely equipped to fuel empathy bias.

The morally deleterious effects of this sort of partisan-based intergroup empathy bias become clear when considering a study by Combs et al. (2009). The study examines the relationship between partisan group identity and the experience of schadenfreude, an emotion defined by the experience of joy when witnessing another's pain. There is a growing literature on the relationship between schadenfreude and empathy, much of which suggests that individuals who empathize more with members of their ingroup, i.e., experience intergroup empathy bias,

38 See Barberá et al. (2015) for a study on "ideological segregation" amongst Twitter users. Interestingly, the study found that users tended to communicate more with those of similar ideological backgrounds when discussing political issues, but that this segregation did not extend to discussion of other current events. In other words, the findings suggest that the extent to which Twitter is an "echo chamber" in which users reinforce their own ideologies may be exacerbated by political issues on which there is already significant disagreement.

are more likely to experience schadenfreude in response to harm to outgroup members.[39] Combs et al. demonstrate schadenfreude based on partisan affiliation in the U.S. They find that Democrats experienced more schadenfreude when reacting to the misfortunes of Republican presidential candidates, while Republicans experienced more schadenfreude when reacting to the misfortunes of Democratic presidential candidates. This result is perhaps not so surprising. However, the study also found that Democrats, especially those with particularly strong party affiliation, experienced a significant amount of schadenfreude in response to reading of an economic downturn that occurred during a Republican administration. The key here is that an economic downturn, though more *politically* problematic for Republicans during a Republican administration, is still a mutual harm for both Democrats and Republicans. This is a case in which partisan group identification causes an ingroup empathy bias in which empathy for the ingroup is so strong that one ignores broader human concerns. It is bad enough that intergroup empathy bias can cause one to ignore the legitimate concerns of those who do not share one's group membership, but it is worse when these concerns are concerns not about divisive values but about the general well-being of all those involved. In sum, Combs et al.'s study offers a glimpse into how favoring empathizing with one's ingroup members can cause one to ignore or have inappropriate reactions to pressing moral concerns that exist outside the confines of intergroup competition.

2.3: Bias of Scope

Intergroup empathy bias is problematic for moral inquiry because it leads us to inappropriately value group membership in our moral decisions. The problem is that we allow an extra factor of group membership into our moral judgments, and that factor should not play a role in our moral decision making. Furthermore, we ignore relevant factors that might be

39 See, for example, Cikara et al. (2014).

appreciated if we were more capable of empathizing with the perspectives of members of perceived outgroups. An empathetic bias of scope is problematic for moral inquiry in a subtly different way. It is problematic because it dampens our capacity to make important considerations regarding the broader impact of our moral decisions and regarding the broader possibilities for alternative moral action. As Bloom notes, the problem is that empathy can lead to innumeracy. In other words, empathy's bias of scope is not problematic because of the implicit addition of an irrelevant factor like group membership to moral inquiry; rather, bias of scope subtracts a relevant factor in moral deliberation, namely consideration of other people affected by our moral behavior and views.

We can see the negative impact of such a subtraction by understanding what Thomas Schelling and others[40] have called the "identified victim effect," our tendency to care much more about problems in which we can identify a specific victim, and to make disproportionate efforts to help that particular victim rather than to help the broader group of affected individuals. This tendency is illustrated by a classic experiment by Batson et al. (1995) in which participants were told that they could move one particular child, Sheri Summers, up to the top of an organ transplant list, but that doing so would mean that Sheri would be moved ahead of different children who may be more deserving (e.g., spent more time waiting for the transplant). The study found that participants did not decide to move Sheri ahead of the other children if given no empathy prompt, but that they tended to move Sheri ahead when given a prompt, similar to the one discussed above in Johnson et al.'s study on race and sentencing, to imagine what Sheri was feeling. In this case, empathizing with one individual led to a moral judgment that was not fair to the other individuals involved. Empathizing with Sheri leaves one blind to the exact same, or

40 See Kogut and Ritov (2005).

even more pressing, needs in other individuals and thus to favor Sheri for a reason that should not be morally relevant to the decision at hand.

In another study, Kogut and Ritov (2005) found that participants were more likely to donate money to fund research into a potentially life-saving drug when they were provided with a picture of a single identified child suffering from the disease that the drug would cure, as opposed to when they were told that there are eight children in need of the drug. Empathy's bias of scope links our moral judgments to factors that are based on our emotional connection to a particular identified victim, but in many cases addressing the problem at hand, in this case a disease affecting more than the one identified victim, needs to involve a more careful consideration of how a broader group of individuals is affected by our judgment.

These studies suggest that empathy's bias of scope is a problem regardless of one's approach to normative ethics. For example, in Batson et al.'s study, we saw that empathy for an individual can cause us to treat other individuals unfairly, to ignore what a deontological theory might hold as their right to equal consideration. On the other hand, in Kogut and Ritov's study we see that empathy for an identified victim can lead us to make moral decisions without paying attention to relevant consequentialist considerations of how many people our action will affect. In either case, we see empathy's bias of scope dampening our ability to recognize relevant moral factors, namely the interests and/or rights of a broad group of affected individuals that should be accounted for in the process of moral inquiry.

2.4: Proximity and Exposure Bias

Thus far we have seen that empathy can lead us to biased consideration of members of ingroups and even to schadenfreude at the pain of members of perceived outgroups. The evidence also suggests that empathy can lead us to biased consideration of particular individuals

over the concerns of larger groups. I now want to consider one final form of bias that is distinct from, yet importantly related to both intergroup empathy bias and empathy's bias of scope; it is a bias towards empathizing with those who are more geographically proximate to us, or to whom we have more media exposure, even when these factors are not morally relevant, and even when they cause us to ignore other morally relevant factors.

In order to empathize, we must have some form of contact with the target of empathy. The issue of a bias of proximity and exposure arises in that geographic proximity and media exposure often dictate which individuals we come into contact with but do so in a way that operates beneath our conscious awareness. I take this to be a morally problematic bias in that our geographic proximity to an individual or the amount of media exposure that an individual receives, are arbitrary factors that should not be directly relevant in moral deliberation in most cases, yet they are factors in determining who our targets of empathy are. The fact that I live nearby a particular hospital does not necessarily make that hospital more deserving of a charitable donation than a hospital in a different country. The fact that a particular political candidate is able to afford more television advertisements does not necessarily mean that he or she is a better candidate than others. Yet because empathy relies on contact, our increased contact with individuals tied to local concerns and our increased contact with individuals who are the subject of greater media exposure can lead us to empathize in a biased manner that favors these individuals.

We can now see how this proximity/exposure bias facilitates empathy's susceptibility to intergroup bias. For example, if one lives in a community that is particularly racially, religiously, or politically homogenous, then most of the individuals one encounters in one's community will be members of one particular race, religion, or political party. If one identifies with that

particular group as an ingroup, then we have a case in which one's geographic proximity facilitates ingroup bias in that the proximity leads to much more contact with ingroups than with outgroups, and, again, it is this contact that leads to empathy.

Geographic proximity or media exposure can also facilitate empathy's bias of scope. We saw this in the case of Sandy Hook discussed in 2.1. The significant amount of media coverage of the shootings led to a substantial amount of donations of all sorts to the victims' families. The families of shooting victims in Chicago do not receive the same amount of media attention. Thus, those who sought to help families of children killed in shootings, certainly a worthy cause, continued to direct their donations to the relatively affluent community of Sandy Hook even when told such donations were not needed; in this case their donations could better help similar families elsewhere. The key is that media coverage plays a role in directing our empathic attention to particular individuals, but level of need is not necessarily tied to level of media exposure. Our empathetic focus on particular individuals can lead us to direct our moral attention away from groups and individuals that are in need but are not the focus of media coverage, or perhaps are too geographically distant for us to have empathetic contact.

While the relationship between empathy's bias towards individuals who receive more media exposure and empathy's bias of scope is perhaps somewhat intuitive, the relationship between empathy's proximity bias and bias of scope is more nuanced. We can see this by examining the findings of a 2018 study by Kogut et al. The study found that while increased distance between a prospective donor and victim decreases the prospective donor's willingness to help, this effect only held in cases in which the victim was unidentified. This suggests that the identified victim effect mitigates empathy's proximity bias. The study's finding is significant in that it provides empirical evidence for the sort of geographic proximity bias discussed above, but

it is also significant in suggesting a way out of this bias, and specifically a way out through utilizing empathy's susceptibility to the identified victim effect, which is precisely the sort of susceptibility at the heart of empathy's bias of scope.

Thus, we have a case in which empathy's propensity towards engaging with specific individual concerns can be utilized as a means of encouraging a *less* biased moral approach. This is a theme that will pervade the remainder of the dissertation, but not at the expense of ignoring the pitfalls of the biases discussed above. In defending the role of empathy in moral inquiry, the key is to define a role which takes advantage of empathy's unique capacity to encourage nuanced individual engagement, but which also limits its susceptibility to do so at the risk of ignoring broader morally relevant concerns, or of adding morally irrelevant factors to the process of moral inquiry.

In this chapter I hope to have shown just how problematic these biased susceptibilities have the potential to be. We have seen that empathy can lead us to consider irrelevant factors of group membership in moral decision-making, to ignore important statistical facts of the scope of our moral behavior, and to inadvertently direct our attention to those whom we happen to have arbitrary proximity or exposure. While these are all significant problems for empathy to overcome, in the following chapters I will argue that they are not impossible to overcome and are not powerful enough to force us to ignore empathy's benefits, specifically empathy's crucial role in critical moral inquiry.

Chapter 3

Constructing Empathetic Emotions, Combating Bias

In Chapter 1, I outlined an account of empathy based on other-oriented perspective taking and affective matching and argued that simulation is a crucial component of how we understand the emotions of others. In Chapter 2, I argued that an unbiased empathetic understanding of others faces serious obstacles, as empirical research shows that we are susceptible to a variety of biases that affect how and with whom we empathize. In the remaining chapters, I will defend a method of moral inquiry that aims to effortfully adjust empathy to correct these biases so as to effectively employ empathy as a means of critical moral self-evaluation, rather than eliminate empathy from the moral life altogether as empathy's critics propose. In order to defend this method of moral inquiry, it is first necessary to outline *how* the process that I have defined as empathy is capable of adjustment. A prerequisite for defending the value of a method of moral inquiry that involves the correction of empathy biases is that these biases are in fact correctable. As such, my goal in this chapter is to defend the view that it is possible to engage in effortful correction of empathy bias.

In order to do so, I will outline and defend Lisa Feldman Barrett's Conceptual Act Theory (CAT) of emotion and relate this theory of emotion to the account of empathy that I described in Chapter 1. CAT is what Barrett calls a "constructionist" theory of emotion according to which emotions are not innate, hardwired reactions to outside stimuli; rather they are multiply realizable situated concepts that are constructed by the agent according to predictive coding in the brain, concepts that are developed and refined over the course of the agent's experience. The upshot of this theory in terms of effortful empathy is that emotion concepts are malleable, nuanced, and dependent on the specifics of each agent's experience, rather than

universal and inflexible. One can develop more fine-grained emotion concepts and can develop a wider variety of emotion concepts by virtue of effortfully seeking out diverse experiences, as situating oneself in these experiences requires that one's emotion concepts adapt and respond to novel challenges. Thus, insofar as one is capable of seeking out new experiences, one is capable of refining one's palette of emotion concepts so as to be better able to construct the emotions of others, and thus to become a less-biased empathizer.

In contrast to what Barrett calls "classical theories of emotion" that emphasize the evolution of automatic, hardwired, reactive neural modules corresponding to particular emotions and to universal experiences and expressions, CAT emphasizes that the neural and experiential architecture of a given emotion varies significantly depending on the context in which that emotion is experienced, and that this experience is a top-down construction based on predictions in the brain rather than a mere reaction to outside stimuli. According to CAT, individuals' experiences and neural realizations of emotions will vary according to the development of their specific emotion concepts, which are based on their specific experience of constructing emotions in similar contexts in the past. Furthermore, emotions, as abstract situated concepts, are multisensory: they incorporate a variety of sense data and contextual information as relevant in the prediction of an emotion experience. The sense data that is deemed more salient in the experience of an emotion will vary according to an individual's particular emotion concept; this difference in salience is a function of the salience of similar sense data in the experience of the emotion in similar contexts in the past.

This account of emotion has important implications for empathy. Because each individual's emotion concept is tied to his or her specific experiences, the best way for one to understand another's emotions is to make the effort to understand the other's specific

experiences, rather than to assume that each individual shares a universal emotional palette and that each individual will signal the same emotions in the same general ways. I will argue that this empathetic effort can come in the form of any of three general approaches. First, it may be more of a long-term, embedded process involving physically situating oneself in a variety of diverse environments so as to have more potential shared experiential reference points to draw on when attempting to construct the emotion of another. I call this the embedded approach. Second, it may be a communicative process involving seeking out direct communication with those whose experiential background is different from one's own so as to better equip one's imagination to construct emotions with others' experiences in mind. I call this the communicative approach, and it is particularly relevant when attempting to understand emotion concepts built on experiences that one cannot directly experience oneself, such as those that may be specific to a member of another race, gender, or generation. Third, empathetic effort may come from imaginative engagement with art, particularly narrative art that depicts novel and nuanced experiences and perspectives. I call this the imaginative approach.

A thread that ties the embedded, communicative, and imaginative approaches together is that they all involve making an effort to broaden one's capacity to engage in the sort of other-oriented perspective taking discussed in Chapter 1. Engaging in any of these three approaches refines an empathizer's emotional capacity because each approach allows the empathizer to accumulate relevant experiential knowledge to incorporate into her existing emotion concepts. As one constructs emotions in novel contexts, one enables one's emotion concept to incorporate those novel contexts as relevant and thus to enable those concepts to become, to use Barrett's language, more "fine-grained." For example, constructing happiness when experiencing the comforts and bonds of enjoying a traditional meal with those from a different cultural

background, whether directly through sharing in such a meal oneself, through communication with those who value such comforts and bonds, or through imaginative engagement with art that depicts such comforts and bonds, allows one to incorporate this experience into a more nuanced concept of what happiness and camaraderie can be and in which contexts these emotions can occur. As one develops this more fine-grained conception of happiness, one is able to empathize with a more diverse array of others' conceptions of happiness, as one can draw on one's own diverse experiences to empathetically construct the happiness of others whose concepts are built on similar experiences.

While emotion concepts are perhaps more directly refined through the embedded approach of engaging with novel experiences, sometimes this process of refining our emotion concepts must involve an effort of the imagination due to limitations on what sorts of experiences we can directly pursue. But we can consciously, actively make this imaginative effort, either through engagement with art that depicts diverse perspectives (the imaginative approach), or through engaging in deep communication with those who have access to experiences that we have not had or cannot have ourselves (the communicative approach). For example, if I want to empathize with a Black person's outrage over the experience of discrimination in the United States, I cannot draw on my own experiences as a white person to simulate a similar emotion of outrage. Furthermore, I cannot seek out ways to directly put myself in the experiences on which the Black person's emotion concept is built. Rather, I must try to communicate with those who have had such experiences or engage with art created by such individuals so as to incorporate insights from this engagement into my own emotion concept of moral outrage.[41]

41 This is not to say that I will or can directly experience the *same* moral outrage as the target of empathy. Empathy occurs in degrees, and there is an important sense in which I will simply not be able to entirely empathize with the

These approaches drive us to engage directly with novel experiences, or to engage indirectly with first-hand accounts or artistic depictions of novel experiences, so as to be better equipped to empathetically reconstruct a *degree* of another's perspective that is built on such experiences. It is important to note that, given that emotion concepts are nuanced and built on individual experience, and that each individual's experiences will at some level be unique, complete overlap of fine-grained emotion concepts is not a realistic goal of effortful empathizing. But I see no reason why this should deter us from seeking the benefits of a higher degree of overlap with the emotion concepts of those who differ from us, the benefits of seeking to refine our emotions in a way that draws on engagement with diverse perspectives and experiences. While we may not be able to achieve complete empathetic understanding of another, empathy only ever occurs in degrees, and degrees of empathy are valuable insofar as they broaden our self-concept and capacity for self-criticism, particularly moral self-criticism, in ways that we cannot achieve by simply ignoring our capacity to engage in any sort of empathetic understanding of others.

Whether the refining of emotion concepts via experience occurs through a direct process of gaining diverse experiences, or through a more imaginative process based on communication with others or engagement with art, the underlying goal is to address instances of experiential blindness that cloud our moral judgment, particularly our critical outlook on our own moral views. Because emotions are constructed via predictive coding, one needs to accumulate a variety of experiences to be able to appropriately predict the emotions that are relevant in a

---

emotional experience of someone from a background that I do not share. I can empathize with moral outrage over racism to a degree, but there remains an important sense in which I cannot experience the emotional response that comes from directly being the target of racism. Nevertheless, the fact that we can empathize to some degree remains important. The goal in terms of moral inquiry is to try to empathize such that we are able to evaluate our own moral perspective from a more distanced, less partial perspective that incorporates the experiences of others. As long as empathy enables us to incorporate those perspectives to some degree, it remains valuable in generating a critical distance that we would not have if we did not empathetically incorporate the perspectives into our evaluation at all.

variety of contexts. In the case of moral problems, the emotions that are relevant are not merely one's own; one must take into account the emotions of others involved. Thus, insofar as empathy is a means of understanding the emotions of others, it is a means of combating experiential blindness in the case of moral problems, and it is a means that we can develop and refine by making a conscious effort to add novel experiences as relevant elements of our emotion concepts.

This leads to the second crucial implication of CAT for empathy: accurately recognizing emotion in others ought not be driven by mere recognition of facial cues, postures, or other physical responses in isolation. These cues are relevant but only when contextualized. According to classical theories of emotion, there are certain universal signals of emotion: wide eyes signify surprise, a smile signifies happiness, a grimace signifies anger. On such a view, these signals can function as cues for us to empathize with the corresponding emotions. A smile may lead us to empathize with happiness, etc. But CAT emphasizes that these signals occur within a certain experiential, developmental, and cultural context that must be considered. For example, in the context of a dispute, a smile or laugh may not signal happiness but rather condescension or even indignation. Tears can signal extreme grief or extreme joy. In some cultures, traditional Western displays of emotion are not considered in terms of emotion at all, but rather merely as behaviors. Thus, when we try to empathize with another, we cannot merely focus on the physical displays of emotion that they exhibit. While we may experience emotional contagion based on such signals, in terms of accurate empathizing it is essential to recognize that this experience is at best underdeveloped due to lack of contextualization, and at worst completely misleading based on our own stereotypes regarding emotion signaling. It is only when we appropriately contextualize our experience of affective matching from the context of the other's perspective that we are able

to achieve some degree of empathy with another person. While emotional contagion is perhaps beyond our control, it is certainly within our control to actively seek the sort of information that can contextualize our experience in order to refine or even reject the initial affective experience. This process of refining through contextualization is what makes for the experience of empathetically constructing an emotion rather than merely experiencing an affective response based on the stimuli of another's bodily signals.

So, understanding emotion in terms of CAT, along with understanding empathy as outlined in Chapter 1, enables us to describe a process of effortful empathy that can combat biases based on the active pursuit of refining emotion concepts through embedded, communicative, or imaginative experiences. In order to make this argument, the chapter proceeds as follows:

In 3.1 I provide a more detailed account of CAT, situating it as a rejection of classical theories of emotion and emphasizing the significance of the multiple realizability, context dependence, empirical development, and predictive character of emotion concepts. I provide empirical evidence from psychology and neuroscience that suggests that emotions are not innate, module-defined, automatic, or primarily reactive, and thus that our emotions can be developed with effort. In 3.2 I discuss the implications of CAT for empathy, elaborating on the points discussed above regarding the embedded, communicative, and imaginative approaches to effortfully developing more fine-grained emotion concepts. My goal is to argue that empathy is a capacity that can be refined with effort and to outline approaches for what that effort looks like. If this argument is successful and the approaches outlined are feasible, then the question as to whether or not we should refine empathy, the question that is largely the focus of Chapters 4-6, remains open for consideration despite evidence of empathy bias.

3.1: Defending the Conceptual Act Theory of Emotion

3.1.1: Classical Theories and "The Paradox of Emotion"

In contrasting classical theories of emotion to CAT, it is helpful to consider what Barrett (2006) calls "the emotion paradox," which she characterizes as follows:

> People believe that they know an emotion when they see it, and as a consequence assume that emotions are discrete events that can be recognized with some degree of accuracy, but scientists have yet to produce a set of clear and consistent criteria for indicating when an emotion is present and when it is not (p. 20).

We can distinguish classical theories of emotion from CAT in terms of approaches to resolving this "paradox."[42] Proponents of classical theories of emotion seek to resolve the paradox by looking for the set of criteria that can identify emotions as discrete events: mechanisms and physiological markers that universally indicate when an emotion such as anger, fear, or happiness is present. In other words, on the classical approach to emotions, insofar as one accepts the paradox, it is that we have not identified such criteria *yet*, but this does not mean that such criteria are not discoverable. Classical theories begin from the assumption that emotions are discrete events definable in terms of universal criteria, then attempt to discover the criteria through various theoretical and empirical approaches to studying emotion.

By contrast, CAT explores the implications of rejecting the initial assumption that emotions are universally definable in terms of specific physiological markers, mechanisms or dedicated neural modules that correspond neatly to discrete emotions in all human beings. CAT rejects the assumption that emotions are natural kinds. In doing so CAT does not reject the possibility of identifying when an emotion is present, but it does reject the idea that one universal

42 Of course, this is not a paradox in the strict sense, but rather an incongruity between how emotions are classically conceived with the findings (or lack thereof) of science. Nevertheless, I will continue to use Barret's language of the "emotion paradox" in framing this problem.

set of criteria will be able to do so. Thus, CAT dissolves the paradox by rejecting its fundamental assumption and in doing so conceptualizes emotions such that it makes sense that we are not able to identify consistent criteria that universally identify emotions. We cannot provide universal criteria for emotions because such universal criteria do not exist.

In this way the CAT response to the emotion paradox is analogous to Wittgenstein's (1953) arguments regarding family resemblances. While we all seem to be able to identify instances of a "game," we are at a loss to provide adequate necessary and sufficient conditions that define the concept of game. In terms of emotion, classical theories continue to operate as if such necessary and sufficient conditions are discoverable for each emotion, whereas CAT is based on embracing the lack of evidence in favor of such conditions and instead focuses on the implications of defining emotions more along the lines of the sort of family resemblance approach that Wittgenstein discusses.

CAT can be summarized in terms of four main postulates: (1) Emotions are multiply realizable; they can be realized by different physiological conditions and behaviors. (2) Emotions are context dependent; the same physiological conditions and behaviors can realize different emotions in different contexts. (3) Emotions are predictive simulations constructed by the agent as she interacts with her environment. (4) Emotions are constructed through an act of categorizing a particular experience according to emotion concepts grounded in prior experience and in a neurological process involving predictive coding.

In the remainder of 3.1, I will defend CAT by considering the empirical evidence in favor of these postulates. The empirical evidence discussed in this section is meant to object to classical theories of emotion by showing the following: (A) Emotions are multiply realizable and context dependent, not type-identical, universal, or innate. They are not expressed by universally

recognized facial expressions, nor are they realized by hardwired autonomic nervous system responses or by specifically dedicated brain circuits. (B) Emotions are not primarily reactive; they involve prediction grounded in the agent's prior experience rather than mere reactions to outside stimuli. (A) and (B) are crucial for understanding our ability to correct empathy biases. First, if emotions are multiply realizable and context dependent rather than type-identical, innate, and universally recognizable, then it is crucial that we understand nuanced considerations of context if we are to understand another person's emotion; we cannot infer emotion based on the presence or absence of universal signals. Second, if emotions are predictions based on prior experience, then we can become better empathizers, better predictors of the emotions of others, by both diversifying and refining our own experiences. This idea will be unpacked in 3.2, but before doing so I will challenge classical theories of emotion and defend the view that emotions are multiply realizable and context dependent and that they are predictions based on prior experience.

3.1.2: Emotions are Multiply Realizable and Context Dependent

According to the classical view of emotion, emotions are not context dependent and they are not multiply realizable. Rather, emotions are universal such that all human beings possess certain emotion circuits that are wired from birth, and when these circuits are activated in any context, we can identify that a particular emotion such as sadness, fear, or anger that corresponds to that circuit is experienced. Furthermore, on the classical view these emotions are expressed by universal physiological signals, particularly facial expressions: a smile indicates happiness, wide eyes indicate fear, a frown indicates sadness. Classical views argue that the recognition of these signals in others as representing certain emotions is linked to the innate and universal repertoire of emotions that all human beings share and that correspond to underlying universal circuits that

trigger such signaling behavior. The reasoning is that, if emotions are universally recognizable in terms of facial expression, even by those with completely different cultural and psychological backgrounds and with no training in facial-based emotion recognition, then these emotions must be innate and universally realized.

Yet, as we will see, there is little evidence that human beings across cultural and environmental contexts do recognize the same emotion in the same facial expressions, particularly without any sort of priming. Furthermore, neuroscientific evidence does not support a type-identity relation between neurophysiological circuits and emotions. The context in which another's facial expression is perceived is crucial to how the perceiver labels the other's emotion, and the context in which a given neurophysiological state is situated is a crucial factor in determining which emotion is constructed. Emotions are recognized and experienced based on contextual information, and the conceptualization of this contextual information will vary depending on an individual's psychological and cultural background. Thus, different facial expressions will be linked to different emotions depending on cultural and environmental context, and the same neurophysiological state may be linked to different emotions when occurring in different contexts. If this is the case, then emotions cannot be identified across age, sex, personality, and culture by means of physical measurements of a person's face, body, and brain; rather, understanding another's emotions requires understanding their emotion concepts and the cultural and environmental context in which a facial expression, bodily signal, or brain state has developed and within which it is immediately situated. Emotions are not innate, and they are not universally recognizable.

Emotion as Identified by Facial Expression

The approach of searching for universal emotion cues in facial expressions has its origins in Charles Darwin's (1872/2005) explanation of emotion in *The Expression of the Emotions in Man and Animals*. Darwin argued that a universal emotion profile is part of human nature. He claimed that human beings evolved to share common abilities to exhibit and recognize emotions based on shared representative facial expressions. Thus, humans do not to need to be trained in any way to recognize that a smiling person is happy, or to smile to express their own happiness; they are simply hard-wired to do so from birth. Beginning in the 1960s, psychologists Silvan S. Tomkins, Caroll E. Izard, and Paul Ekman[43] aimed to put this idea to the test through designing an experimental method in which participants are tasked with identifying the emotions displayed by actors portraying facial expressions deemed to be representative examples of displays of six basic emotions: anger, fear, disgust, surprise, sadness, and happiness.

This approach, known as the basic emotion approach, to testing the universality of emotion recognition is still used in contemporary research on emotions. A standard experiment using this method will show participants one of the facial expressions (e.g., a face with a wide-eyed expression, meant to be representative of fear) alongside a list of emotion words that could possibly describe the emotion being expressed (sadness, anger, disgust, anger, fear, or happiness). The participant is then asked to choose the emotion word that best describes the facial expression. Participants are said to have accurately identified an emotion if they label the facial expression with the emotion word that the researchers take to be represented by the facial expression. Another sort of experiment provides participants with two facial expressions and a story explaining an experience of emotion (e.g., "her mother just died, and she is very sad"). The participant must then choose the expression that best fits the story.[44]

43 See Ekman et al. (1969), Izard (1971), and Tomkins and McCarter (1964).
44 This method was introduced by John Dashiell (1927) and utilized, for example, in Ekman and Friesen (1971).

The aim of these kind of experiments is to test how consistently individuals can label the expression with the emotion word that the researchers have deemed ought to correspond to that expression. If participants, particularly across cultural variations, consistently label the facial expressions with the same "accurate" emotion word in this sense, then the researchers take it that this is evidence of the universality of emotion recognition and thus of the innate character of the basic emotions listed above. As it turns out, many studies in the ensuing years have found that participants across a variety of cultures do accurately recognize the facial expressions in basic emotion tests according to researcher expectations. Barrett (2018) notes that a review of cross-cultural studies on emotion recognition based on facial expression conducted by Russell (1994) found that,

> Test subjects from all around the world (Germany, France, Italy, United Kingdom, Scotland, Switzerland, Sweden, Greece, Estonia, Argentina, Brazil, and Chile) choose the expected word or face about 85 percent of the time on average. In cultures that are less like the United States, such as Japan, Malaysia, Ethiopia, China, Sumatra, and Turkey, subjects match faces and words slightly less well, responding as expected about 72 percent of the time. (p. 44)

The drop in accuracy as cultural variation is amplified should already be cause for reconsideration of conclusions of universality, but 72 percent accuracy is still a significantly high percentage. However, more recent research has called into question just what is shown by these basic emotion experiments. Two kinds of research are particularly relevant. The first is research that removes or alters contextual information and finds a significant drop in success in terms of participant ability to "accurately" label basic emotion facial expressions according to the posits of the basic emotion method. The second is research that measures physiological responses in the

bodies, brains, and faces of individuals as they experience emotions and finds no type-identity relation between physiological factors and instances of emotion. Let us consider each in turn.

Widen et al. (2011) altered the standard basic emotion method by removing the list of words as options for labeling a given facial expression, instead asking participants to engage in "free labeling" in which they entered a label for the expression by drawing on their own unique repertoire of emotion concepts rather than on a list of six stipulated basic emotions. In this study participants used the expected emotion label (or a synonym) only 58 percent of the time—not the sort of result that supports claims of universality. One potential explanation of this result is that providing a list of emotion words for participants to choose from primes participants to simulate the emotions listed in order to consider which emotion word is the best fit. In doing so, they will choose the expected word, as simulating the emotion that corresponds to that word best fits with the face in comparison to the other options. In other words, the list of potential emotion labels provides a context in which the participant is more likely to associate the facial expression in question with the expected label, particularly because the expected label is seen as the best choice out of a forced choice of a limited number of options. By contrast, when subjects are provided with only a face and no contextual clues as to what the "correct" label may be in terms of a list of possible choices, subjects will not label the faces according to researcher expectations as often, as they can draw from dozens or hundreds of emotion concepts based on their own experiences with similar faces in a variety of contexts. No list primes them to contextualize the face relative to any specific set of emotions, so the results show much more variety in terms of which emotions are recognized. The idea is that listing a limited set of emotion words provides a forced choice that primes the participants to identify facial expressions in a certain way. This is further demonstrated by a study by Lindquist et al. (2006), in which no emotion words were

provided and participants were asked to determine whether two separate faces expressed the same emotion. Participant answers matched the expected results as stipulated by the basic emotion method only 42 percent of the time.

This effect is even more pronounced in non-Western cultures. For example, Gendron et al. (2014a, 2014b) found that when members of the Himba, a remote culture located in northern Namibia, were asked to freely label images of the facial expressions standardly used to signify basic emotions in Western research, their labeling behavior did not correspond to labeling behaviors of participants from Western societies. As Barrett writes, "smiling faces were not 'happy' (*ohange*) but laughing, wide-eyed faces were not 'fearful' (*okutira*) but 'looking' (*tarera*). In other words, the Himba categorized facial movements as behaviors rather than inferring mental states or feelings" (2018, p. 49).[45]

Barrett notes that meta-analyses[46] of the original cross-cultural facial recognition studies such as those conducted with the Himba and other non-Western cultures reveal that,

> [o]f the seven samples using test subjects from remote cultures, the four that used the basic emotion method provided strong evidence for universality, but the remaining three used free labeling and did not show evidence of universality" (2018, p. 52).

CAT can provide an explanation for this result: the samples that employed the basic emotion method of including a list of words primed subjects to simulate those emotion concepts and thus

45 This is not necessarily to say that the Himba lacked concepts for "happy" or "fearful." Rather, such results show that certain expressions do not necessarily signify such emotions and that other cultures' conceptualizations of emotions such as happiness or fear may be different from our own—these emotions are not universal. The concepts for "happy" and "fearful" involve different contextual information for the Himba, thus we cannot assume that they are unable to empathize with Westerners based on these sort of results, but rather should understand that their emotion concepts do not make facial expressions salient in the same way in which Western emotion concepts often do. Rather, their emotion concepts are more focused on actions. This sort of action-based focus is also found in certain Japanese emotion concepts and in the Ifaluk people of Micronesia (Lutz, 1983). It would be a mistake to try to empathize with a member of the Himba or the Ifaluk based on recognition of their facial expressions; it is far more important to learn about their cultural beliefs and practices so as to recognize their different ways of expressing and recognizing emotions.

46 See Russell (1994), Gendron (2014b)

to contextualize the face in question as relating to the concept simulated. As Barrett writes, "the basic emotion method guides people to construct perceptions of Western-style emotions" (2018, p. 52). In the absence of such linguistic guidance, participants do not construct emotions that conform to the expectations of the basic emotion method, but rather construct emotions based on their cultural background. It is not clear how the classical method can explain the gap between the basic emotion method and free labeling results. If recognition of emotion expressions is universal, then it should not matter whether a participant is given a forced choice of possible labels; the participant should simply label the expression according to the universally understood emotion that it represents. Free labeling data shows that this is not the case.

So, as Barrett et al. (2011) write, "words constitute a clear example of a perceiver-based context because they provide a top-down constraint in emotion perception, contributing information over and above the affective meaning available in structural information of a face" (p. 287). But it is not only emotion words that provide contextual clues that influence the labeling of facial recognition. Environmental and biographical factors are important as well, as discussed in the introduction to this chapter. A smile can indicate warmth in the context of an encounter with a friend or condescension in the context of a meeting with a boss who just excoriated your latest report. A furrowed brow can indicate exasperated frustration or calm concentration when solving a math problem depending on if the potential solver is an experienced mathematician or not. The emotion one would perceive based only on an image of Serena Williams' face after winning a tennis match given no other contextual information (e.g., pain, anger, frustration), is likely much different than the emotions one would perceive based on

the image of that same facial expression contextualized in terms of her bodily posture and the contextual environment of the tennis match.[47]

In addition, work by Aviezer et al. (2008, 2012) shows that bodily actions play a key role in emotion labeling. In the studies, participations examined edited images in which stereotypical faces associated with the basic emotion method were grafted onto bodies that did not match the expected emotion—for example, a face associated with anger could be grafted on to a body holding a dirty diaper. Participants tended to identify the emotion that would match the body rather than the face suggested by the basic emotion method (e.g., an angry face on a body holding a dirty diaper would be labeled as "disgusted" rather than "angry").

Historical research on emotion further supports a lack of universality based on facial expression. As Beard (2014, p. 75) argues, ancient Romans did not smile to express happiness. There is in fact no word for "smile" in Latin. This does not mean that ancient Romans never made the expression that we call a smile. It just may mean that such an expression did not have the emotional significance that we ascribe to it in the present day. In other words, perhaps the smile is not an innate and historically universal signifier of human happiness.

The connection that all of this research shares is that it suggests that a facial expression alone simply does not provide enough information to indicate an instance of any emotion, and that when we are able to recognize emotions it is due to a variety of contextual factors that are particular to the case at hand and to our own experience with prior cases. Emotions are not universally expressed or recognized in terms of facial expressions. These expressions signal emotions based on their situation within a given context, and the ways in which contextual clues are interpreted are based on each individual's particular culturally and environmentally influenced emotion concepts.

47 This example is drawn from Barrett et al. (2011).

Emotions are not Neurophysiological Fingerprints

Other researchers pursuing the classical method have focused on locating universality in the neurophysiological states of people as they experience an emotion. Whereas the basic emotion method involved human judgment of others' emotions based on facial expression, this method, which I will call the neurophysiological method, is focused on precisely measuring bodily changes, muscular movements, and brain activity in the subject who is experiencing an emotion. The goal is to identify shared neurophysiological networks that realize emotions such as sadness, fear, or surprise in all human beings regardless of context. This sort of identification would support the Darwinian classical emotion thesis of universally shared emotion circuits hard-wired at birth.

But the empirical evidence does not support this sort of type-identity relation between neurophysiological states and emotions. For example, one approach of the neurophysiological method is to use facial electromyography (EMG) to measure the facial movements of subjects as they experience an emotion based on electrical signals generated by facial muscles as they move. EMG technology is used to measure subtle, perhaps even imperceptible facial movements that could indicate the presence of an emotion.[48] If it turns out that subjects consistently display a certain type of facial movement when experiencing a particular emotion and this sort of facial movement is unique to that emotion, this will count as evidence of a universal physiological type-identity relation for that emotion. This is not what such studies find. Although EMG measurements do seem to show a correlation between certain facial movements and the affective experience of pleasantness, as well as a correlation between certain facial movements and the affective experience of unpleasantness, they do not reveal a type-identity relation between

48 See Tassinary and Cacioppo (1992)

specific facial movements and specific emotions such as anger, sadness, etc.[49] Recall from Chapter 1 that there is an important distinction between affective valence and emotion. While the affective experience of pleasant or unpleasant feelings is an important part of constructing an emotion, it is not sufficient for the experience of emotion, as emotions are defined in terms of contextualized affective experience; we cannot tell what emotion is being experienced just based on the fact that it is pleasant or that it is unpleasant. Thus, while perhaps EMG can tell us something about the universality of affective expression, this does not tell us anything significant about the universality of specific emotions. Perhaps facial expressions can help us distinguish between which emotions are pleasant and which are not pleasant, but this is a different sort of process than distinguishing between specific pleasant emotions or between specific unpleasant emotions. EMG cannot help us make such specific distinctions without consideration of context. As such, EMG testing is unable to establish the universality of even the basic emotions stipulated by the basic emotion method.

The issues encountered by both the basic emotion method and neurophysiological methods based on EMG measurements of facial movements suggest that looking for universality of emotions in facial expressions is the wrong route to take. But this does not necessarily rule out other candidates for emotion fingerprints. Research has also focused on identifying type-identity relations between emotions and physiological responses in the autonomic nervous system (ANS),[50] as well as between emotions and neural modules in the brain. While the evidence does not show that emotions are realized by innate, universally shared facial expressions, perhaps emotions are realized by innate, universal physiological responses or neural architectures that have evolved to produce similar emotions in all human beings.

---

49 See Larsen et al. (2008).

50 Autonomic nervous system responses include, for example, heart rate, blood pressure, and skin conductance.

For example, in an influential study, Ekman et al. (1983) measured variations in heart rate, temperature, skin conductance, and arm tension during the experience of emotion. The experimenters sought to evoke the basic emotions of anger, sadness, fear, disgust, surprise, and happiness, with the goal of correlating specific measurements of bodily states with specific emotions. In order to evoke these emotions, Ekman et al. had participants hold a facial expression associated with a basic emotion in accordance with the expectations of the basic emotion method (e.g., holding a frown was meant to evoke sadness, holding a smile was meant to evoke happiness). Participants could use a mirror to examine their own facial expressions. The study found that holding these facial expressions changed participants' autonomic responses in specific ways. For example, participants' heart rates were faster when holding a scowl (meant to evoke anger) than when holding a smile (meant to evoke happiness). Such results were meant to show universal autonomic characteristics of emotions.

Barrett offers a different explanation. She notes that the participants for this study shared a background in Western culture that already associates these expressions with particular emotions.[51] So, when asked to frown, for example, they could anticipate that the researchers were attempting to evoke sadness, and Barsalou et al. (2003) has since shown that this sort of conceptual understanding can lead one to produce the heart rate and other physical signals measured by Ekman et al. If it is this background, culture-based conceptual understanding and not the actual act of frowning that causes bodily changes, then we would expect that non-Western cultures would not display the same bodily changes when given Ekman et al.'s test. This is precisely what a study of the Indonesian Minangkabau people of West Sumatra conducted by Levenson et al. (1992) found. The Minangkabau did not have the background understanding of Western emotions that Ekman et al.'s Western participants had, and they did

51 See Study 4 in Levenson et al. (1990).

not experience the same ANS responses found by Ekman et al. In addition, Minangkabau participants reported feeling the expected emotion associated with the corresponding facial expressions stipulated by the basic emotion method much less frequently than did Ekman et al.'s subjects.

More broadly, in a meta-analysis of emotion research focused on ANS responses that examined 202 studies, Siegel et al. (2018) found that, although these studies demonstrate a general increase in ANS activity during the experience of emotion, the pattern of this activity does not distinguish one emotion category from another. Furthermore, the meta-analysis found ANS variation accounted for a significant amount of overall variation within emotion categories across the studies, meaning that different studies found different ANS patterns for the same emotion. Lastly, the method used to evoke emotion did not account for variability of ANS response within an emotion category, meaning that different studies that used the same method to invoke emotion (e.g., film, imagery, facial expressions) produced varying results regarding the ANS activity corresponding to a particular emotion category. CAT is better equipped to explain Siegel et al.'s results than are classical theories. According to CAT, emotions are multiply realizable, meaning that it is not surprising that different studies found different sorts of ANS activity corresponding to the same emotion category.

So, at this point, it seems that the empirical evidence does not suggest a type-identity relation between specific emotions and ANS responses, nor does it suggest a type-identity relation between emotion and facial expressions. Thus, a proponent of emotions as innate, universal biological fingerprints must turn elsewhere: there remains the possibility that emotions have type-identity relations to specific neural modules.

Much of the research focused on defining emotions in such a way involves patients with brain lesions in certain areas of the brain thought to be universal, hard-wired loci of emotions. For example, the amygdala is often proposed as the locus of fear, and this hypothesis is supported by research on patients such as SM, who as we saw in Chapter 1 has significant damage to her amygdala and is also deficient in the ability to both experience fear and to perceive fear in others. In Chapter 1, I argued that this paired deficit in emotion experience and emotion recognition provided important evidence in favor of simulation-based theories in ToM. In terms of research on the potential universal neural localization of emotion, the neurophysiological aspect of SM's condition is emphasized. The argument is that the correlation between SM's amygdala damage and her deficits in fear, and only fear, speak to a specific dependence of fear on the neurons of the amygdala. This sort of argument concludes that fear can be discretely defined in terms of a neural module in the amygdala, as generally proposed by classical emotion theories.[52]

However, a separate study[53] of identical twins who both suffered similar amygdala damage to SM due to the same rare disease (Urbach-Wiethe disease) found that though one twin, BG, has similar fear deficiencies to SM, the other twin, AM, has normal experiences and perceptions of fear. The takeaway of such a study is that the amygdala is not necessary for generating standard fear responses; fear is multiply realizable, even within the brain. While the results regarding SM and BG suggest that perhaps the amygdala is often importantly involved in constructing and instantiating emotion concepts related to fear, the results regarding AM suggest

52 Though I do not take cases such as SM's to demonstrate the existence of emotion fingerprints in the brain, it is important to note that this in no way discounts the force of SM's case in providing evidence for simulation-based explanations of ToM. The crucial fact regarding simulation theory is SM's paired deficit of emotion *experience* and recognition, and this deficit of experience retains the same significance vis-à-vis simulation in ToM regardless of what neural mechanisms underlie it.
53 Becker et al. (2012).

that it is also possible to bootstrap other areas of the brain into providing functions associated with normal fear response. As such, fear is not reducible to amygdala activity.

In addition to evidence based on brain lesions, researchers have looked to fMRI scans in order to identify dedicated neural networks that correspond to particular emotions. If we can show consistent activity in a particular network of neurons that corresponds to a participants' experience of some emotion, and only that emotion, then we can point to that specific network of neurons as the locus of that emotion. However, this idea is contradicted by a meta-analysis of hundreds of neuroimaging studies on emotion from 1990-2007, conducted by Lindquist et al. (2012). Lindquist et al.'s analysis was based on dividing the brain into small three-dimensional sections, voxels, and tracking the activity in these voxels across neuroimaging studies of emotion. When the probability of activation for a certain voxel was greater than chance during the perception or experience of an emotion in a given study, this was counted as statistically significant, and when this significance held across a significant number of studies, the activity was counted as consistent. A classical view of emotion, which Lindquist et al. refer to as a "locationist" model, would predict that certain combinations of voxels would be significantly and consistently activated across experiences or perceptions of specific emotion categories such as fear, sadness, anger, happiness, surprise, and disgust. Importantly, such significant and consistent activation would have to hold *only* for discrete emotion categories; if the activation was consistent across multiple emotion categories, then this is evidence of the context dependent role of the network of voxels in generating emotions, given that the network would sometimes be active in generating one emotion and sometimes active in generating another emotion. The meta-analysis found that no specific brain regions are consistently and specifically activated across instances of a single emotion category. For example, voxels associated with the amygdala were

found to be consistently active across instances of fear, but they were also found to be consistently active in instances of anger, disgust, sadness, and happiness.

This sort of evidence speaks to two key features of the brain that Barrett argues suggest that emotions are multiply realizable and context dependent: degeneracy and core systems. Degeneracy means that many different combinations of neurons can produce the same outcome, including an emotion. A core system in the brain is one network that participates in generating a variety of mental states, including multiple emotions. Both of these concepts are opposed to the classical emotion thesis that a single network of neurons or a single part of the brain is dedicated to producing one and only one emotion.

3.1.2: Emotions are Predictions Coded by Experience

So, the evidence does not support classical views of emotions as universal, innate, and type-identical to physiological responses. Emotions are not definable by facial expression, by ANS response, or by neural modules. As such, we are left with the question of just what exactly emotions are. In what follows, I will defend the CAT answer to this question: emotions are situated multisensory predictions coded by prior experience.

Understanding emotions as predictions is consistent with and builds upon the significant amount of research emphasizing the role of predictive coding in other areas of experience, including visual perception, auditory perception, and proprioception.[54] The takeaway from this research is that conscious experience is in large part a function of predictions of potential experiences; these predictions are then modulated by outside stimuli in the case of prediction errors. The correction of prediction errors cures what Barrett calls "experiential blindness," and

54 See Clark (2013) for an extensive discussion of the development of predictive coding explanations of perception and its broader implications for the nature of conscious experience. See also Bar (2007, 2009) and Barsalou (2009).

when this blindness is cured, we are better able to make predictions in similar contexts in the future.

The fundamental point underlying CAT is that emotions are predictions in the same way that other experiences are predictions. In the same way that previous visual experiences give meaning to present visual experience by influencing which experiences are predicted in the present, previous emotional experiences give meaning to present emotional experiences by predicting which emotions are experienced in the present. The act of making a prediction based on emotion concepts is the act of constructing an instance of the experience of that emotion.

We can think of emotions as hypotheses based on a collection of experiential evidence. As hypotheses, emotion predictions are open to revision given even subtle shifts in experiential context. For example, suppose that you see a certain brown, elongated shape while hiking in the woods. Based on your prior experience, you experience a surge of fear as you believe the shape is a snake. However, as you begin to run you get a closer look at the shape and realize it is only a fallen branch. Your feelings of fear subside, and you experience a sense of relief as you begin to feel your heart rate decrease. In such a case your brain predicted that the shape was a snake and constructed an instance of fear according to the role of snakes in your emotion concept of fear. This particular instance of the concept involved an increase in heart rate and adrenaline and the goal of escaping from danger. However, you were able to correct the prediction error of this construction by shifting your visual angle. In doing so, you furnish your emotion concept with additional information that will be relevant on future hikes through the woods. Your brain can now incorporate this experience into its predictions, affecting whether constructing an instance of fear will be your response to seeing similar sorts of shapes in the context of a walk in the woods. This experience has factored into curing a certain experiential blindness: a lack of

experience of falsely identifying snakes based on certain visual properties. Given such an experience, you may construct fear differently or not at all in future contexts that share sensory similarities. Or you may not if you are particularly afraid of snakes, as a single experience may not be significant enough to alter your entrenched phobia. In any case, the key is that the experience of an emotion is a prediction, but it is revisable. Furthermore, this act of revision, of correcting prediction error in present experience, can alter your emotion concept, which then informs how you will experience emotions in the future. Both predictions and the correction of prediction errors influence your experience of emotion. As you cure certain areas of experiential blindness, your emotion concepts become more refined and your predictions become more nuanced, but given that each context in which an emotion is constructed has the potential to be subtly different than those involved in the existing emotion concept, the possibility of further revision of the concept through correction of prediction error remains open.

It is also crucial to note that emotion concepts are abstract and multisensory,[55] meaning that they incorporate a variety of situated sensory information such as visual perception, auditory perception, proprioception, and interoception (the perception of internal bodily changes). The instance of an emotion that is constructed will be based on a combination of the array of sensations experienced in a particular context. Thus, emotion concepts incorporate non-emotional sensory concepts (e.g., "snake," "woods," "threat"), with each instance of an emotion involving concepts that may not be shared by another instance. Constructing fear upon seeing what you take to be a snake involves not only a certain visual awareness of your surroundings and the elongated shape, but also an interoceptive awareness of your bodily states (e.g., heart rate, sweat, etc.), and auditory awareness (rustling sounds, screams of companions, etc.). A subtle shift in any of these contextual factors may alter the construction of emotion, depending

55 See Wilson-Mendhenhall et al. (2010).

on how salient such factors are in one's emotion concept. Furthermore, these sensory concepts will not necessarily be involved in the construction of fear in another context. Seeing the same branch in a fireplace will not be involved in a construction of fear, unless perhaps you have a phobia of fire. Crucially, in the case of such a phobia, the experience of fear constructed would not be the same as that constructed in the context of the woods. Fear of fire and fear of a snake are different instances of the emotion of fear, with potentially different physiological and phenomenological profiles. This is not problematic but rather is to be expected based on the multiple realizability and context dependency of emotions.

The last feature of constructed emotions that will be relevant for the discussion of effortful correction of empathy bias is that emotions are predictions based on goals. The brain is constantly making predictions, but the prediction that ultimately wins out and becomes your experience is that which fits a particular goal in the context in which you are situated. This is not to say that we construct an ideal world of perceptions in which we perceive only what we would like to perceive. Our actions are ultimately accountable to the underlying goal of survival, which of course involves being adaptable to the actual social and environmental conditions in which we find ourselves, rather than being blissfully but dangerously ignorant of such factors. Thus, the goal of perception is to accurately predict the environmental context in which one finds oneself such that one is able to productively engage with the environment; this is why we correct prediction errors.

But our goals are not only about survival. Each individual has unique goals, and these goals play a role in an individual's emotion concepts and thus in how a particular instance of an emotion is experienced. For example, suppose I am angry at a colleague for failing to complete his portion of a shared project on time. There are a variety of instances of anger that I could

construct in such a situation: I may yell at my colleague, I may vent to a friend, I may seethe quietly at my desk. Each of these options is a part of my emotion concept of anger, but each has different phenomenological and physiological profiles. The instance of anger that I construct will ultimately be that which fits certain goals in dealing with the colleague. Do I want to embarrass him (yelling)? Do I simply want to destress (venting)? Do I want to avoid confrontation (seething)? This may not be a conscious process in which I choose which emotion I feel based on my favored goal. Rather, my brain makes an array of predictions of potential emotional responses based on the context and my experience of the consequences of different emotional responses in similar situations. If yelling at those I am angry with has been rewarding in some sense in the past, this prediction may win out, but it also may not, as there is always the possibility of prediction error. Perhaps I lash out my colleague in anger but feel terrible after or lose my job. These consequences will impact my emotion concept of anger and will play a role in how and when I construct instances of anger in similar contexts in the future.[56]

It is important to note that, in predicting emotional responses based on goals, it is not necessary that we always actually arrive at what we ultimately consider the best emotional response. A sort of uncontrolled and ultimately undesired anger may often be the instance of anger that we experience, rather than the sort that may be most productive to our long-term goals or desires. However, this sort of uncontrolled anger is still tied to *some* goal. There is a sense in which we do desire to lash out and in which it feels good to angrily vent. Different instances of an emotion may be tied to different goals, e.g., the goal of long-term career success vs. the goal of feeling the rush of lashing out. While it would be ideal if the brain always constructed the instance of emotion that was most in line with what we would want to call our "true" self, or our

56 Of course, it is not always ideal to refine one's emotion concepts in such a trial-and-error manner when the consequences of error are significant. As we will see in 3.2, this is one reason why imagination in general and empathetic imagination in particular play an important role in developing more fine-grained emotions.

"true" desires, the brain is just not a perfect predictor in this way; different goals compete with one another, and some instances of emotions that satisfy goals we may ultimately reject upon calm collected reflection will win out. Analogous examples include the emotion predictions that cause an addict to continually use drugs, or that cause someone to experience irrational phobias. In the case of the addict, ultimately the instance of desperation that leads to satisfying a craving may win out over the instance of desperation, or of some other emotion such as regret, that would lead one to not use or to check into rehab. In the case of irrational phobias, the instance of fear that leads one to the comfort of avoiding the object of fear may win out over an instance of frustration that leads one to confront it. The point is that, of course emotional experiences are not always constructed to accomplish our ideal goals, but this does not mean they are not goal-directed. It merely means that we often have competing goals, and the brain is far from perfect in predicting which goals are ultimately the best. Nevertheless, as we will see in 3.2, because emotions are refinable through experience, we can take active steps to more appropriately calibrate our emotion predictions to our favored goals.

So, according to CAT, instances of emotions are predictions that are constructed according to emotion concepts that are (1) experience-based, (2) revisable, (3) multisensory, and (4) goal-directed. We ought to keep these aspects of constructed emotions in mind as we examine how it is possible to correct empathy bias with the embedded, communicative, and imaginative approaches. However, before doing so I want to briefly highlight the empirical evidence in favor of this account of the predictive character of emotions. This evidence suggests that the predictive model of emotion favored by CAT is not an arm-chair theoretical explanation of emotion, but rather is an empirically supported view based on the structure and function of the brain.

It makes evolutionary sense that we would evolve a brain that operates through prediction. In terms of both energy and time it is inefficient to compute a continuous stream of discrete perceptions from scratch in each moment of experience, just as it is inefficient for a program to compute each pixel of an image in each frame of a digital video. This sort of approach unnecessarily wastes energy and time on reconstructing redundant features. In a world in which there is a significant amount of consistency from moment to moment (or in a video file in which there is a significant amount of consistency from frame to frame) it makes more sense to predict a certain level of regularity and adjust the prediction as features change, rather than continuously detect each aspect of an experience in each moment.[57]

It appears that the brain has evolved just such a strategy. For example, only a small fraction of the neural connections in area V1 of the visual cortex, which is responsible for mapping the visible world, carry input from the eye to the cortex.[58] The vast majority of connections provide predictive information from other parts of the visual cortex. In a more reactive visual system one would expect that the majority of connections would provide input from the eye, as the eye is responsible for detecting the input of light waves from the outside world. This gap between the resources that the brain devotes to prediction and to input from the outside world is not unique to vision; it can be found across sensory modalities.[59]

In terms of emotion, prediction involves what Barrett calls a "cascade" of concepts, as each instance of an emotion will involve multisensory predictions involving externally-oriented sensations such as vision, hearing, touch, etc., as well as the interoception of bodily changes in

57 See Raichle (2010) for an in-depth discussion of the metabolic cost and efficiency of a predictive brain versus that of a "reflexive" brain. Raichle argues that the majority of the brain's metabolic energy is devoted to intrinsic networks involved in prediction, rather than to networks focused on sensory input.
58 See Olshausen and Field (2005).
59 For example, Shipp et al. (2013) find evidence of predictive coding in the motor cortex, and Barrett and Simmons (2015) show the role of predictive coding in interoception. See Friston (2010) for a general model of sensation as prediction.

heart rate, adrenaline levels, etc. The key is that emotions involve the brain's intrinsic networks, networks of neurons that are active with no external catalyst. In particular, researchers have found that the default mode network, which is part of the interoceptive network responsible for predicting experience based on internal changes in one's body, is active across the experience of emotions such as anger, disgust, surprise, happiness, sadness, and fear.[60] The default mode network is associated with conceptual prediction, but can be distinguished from areas associated with specific sensory activity such as motor or visual activity.[61] Neurons in this network do not directly make specific sensory predictions but rather "represent the highest-level, efficient, multisensory summary of the instance" (Barrett, 2018, p. 311) of a concept. Thus, the consistent activation of the default mode network across emotions speaks to the role of conceptual prediction of the sort I have been describing in the experience of an emotion. We can think of emotions as predictions launched in the default mode network that inform specific sensory predictions in different areas of the brain;[62] these predictions are then adjusted based on sensory inputs. As emotional experience is adjusted based on prediction errors, our emotion concept is revised, which may lead to different sensory predictions in future instances.

Importantly, there is not one discrete pattern of activity within the default network that is consistent across instances of the same emotion; emotion concepts are not located in one pattern of neurons within the network. Rather, many different patterns of neurons within the network can be involved in the construction of the same emotion in different instances, and the same pattern of neurons can be active in the construction of different emotions. In other words, while the consistent activity of the default network across emotions speaks to the role of prediction in the experience of emotion, the lack of consistent activity within the network corresponding to any

60 See Kober et al. (2008) and Lindquist et al. (2012).
61 See Binder et al. (1999, 2009) and Spunt et al. (2010).
62 For neuroimaging evidence that supports this account, see Wilson-Mendenhall et al. (2010, 2013, and 2015).

one emotion category speaks to the multiple realizability and context dependence discussed in 3.1.1.

We can now see that empirical evidence supports the four fundamental postulates of CAT, as well as the CAT approach to dissolving the paradox of emotion. Emotions are: (1) multiply realizable rather than locatable in dedicated brain modules or neurophysiological responses, (2) context dependent rather than realized by the same behaviors and neurophysiological processes in every instance, (3) predictions modulated by the environment rather than reactions triggered by the environment, and (4) coded in the brain based on prior experience rather than hard-wired at birth. This antiessentialist, constructionist model of emotion explains the emotion paradox: although we are able to experience and recognize many emotions as a result of our particular past experiences of a given cultural and psychological history, we cannot provide necessary and sufficient conditions to define emotions, because such conditions do not exist.

We are now in a position to understand how conceptualizing emotion in terms of CAT relates to the effortful correction of empathy bias according to the three approaches introduced at the beginning of the chapter: embedded, communicative, and imaginative.

3.2: Effortful Empathy and Constructed Emotion

In the previous section we saw that emotional experience involves the correction of prediction errors caused by experiential blindness. The key to effortful correction of empathy bias is that experiential blindness is a function of a lack of experience, and it is within our power to effortfully compensate for this lack. Without a diverse variety of experiences, one's emotion concepts will not be fine-grained enough to adapt to diverse conditions. In terms of empathy, one will be unable to adapt to diverse perspectives; one will experience empathy bias. Emotions,

including empathetic emotions, are constructed, and the tools involved in this construction are emotion concepts built from prior experience. As such, the less experiential background that we share with one another, the less equipped we will be to empathetically construct emotions that are similar to the emotions of others.

A connection between a gap in prior experience and a gap in ability to empathize provides an explanation for intergroup empathy bias; we will be biased towards empathizing with those who share our experiential backgrounds because our emotion concepts will be most similar to those people. It is easier to construct the emotions of those whose experiential backgrounds are similar to our own, and these individuals will tend to come from perceived ingroups. But this does not mean that it is impossible to construct the emotions of those who are not part of an ingroup, or who have different experiential backgrounds than our own. Because emotions are not hard-wired innate reactions, our emotions are within our power to refine. Furthermore, as we saw in Chapter 1, empathy is not merely the sharing of affective valence, but rather involves the process of appropriately contextualizing affective experience. CAT tells us that the ability to engage in this contextualization is built up from experience. We refine our ability to construct emotions based on experience, and this process is never complete; our emotion concepts can always become more wide-ranging and fine-grained as we engage with different situations to which we were previously experientially blind. As we develop this emotional depth and granularity, we can better contextualize the affective experience of others, because our own emotion concepts are better attuned to the variety and nuance of relevant contextual factors that may be involved in another's emotion concepts. In other words, developing emotional depth and granularity via experience improves our ability to empathize.

Thus, the correction of empathy bias can be thought of as a correction of experiential blindness. We can correct this blindness through a conscious effort to develop more wide-ranging and fine-grained emotion concepts that overlap to some degree with the emotion concepts of others. The more fine-grained our emotion concepts are, the more flexible we will be in terms of being able to construct instances of an emotion from another's perspective. Effortful correction of empathy bias involves effortful correction of experiential blindness. This correction lessens prediction errors that occur when trying to empathetically construct the emotions of others.

In this section I will argue that there are three general approaches we can take to correcting the experiential blindness that underlies empathy bias. The first is what I will call the embedded approach. This approach involves embedding oneself in unfamiliar situations so as to enable one's emotion concepts to adapt to novel environments. Correcting experiential blindness through direct exposure to novel contexts leaves one's emotion concepts more fine-grained and thus leaves one better equipped to simulate emotions from the perspectives of those who share similar experiences.[63] The second approach is what I will call the communicative approach. This approach involves direct communication with others with the goal of drawing on such communication to develop more fine-grained emotions relating to experiences with which one is unfamiliar. The communicative approach may take the form of a group discussion or a one-on-one dialogue with those who have had experiences that one has not had oneself, or that one cannot or should not seek out directly. The third approach is what I will call the imaginative approach. This approach involves engagement with narrative art that portrays novel situations and perspectives that are perhaps not readily available in one's day-to-day life, and that provide a

63 As we will see in Chapter 6, this is the approach favored by Jane Addams in *Democracy and Social Ethics*, though she did not express her views in terms of constructionist theories of emotion but more in terms of the general benefit of conscious effort to place oneself in novel experiences in order to understand different value systems..

degree of exploration and focus that is unique to an artistic treatment of emotions and moral problems.

These three approaches share a common goal: refining empathetic capacity through experience. In terms of moral inquiry, refining empathetic capacity means becoming more attuned to emotional evidence that is relevant when addressing moral problems. One becomes better able to consider the emotional perspective of others on a given moral issue, because one has made a conscious effort to engage with the experiences that shape those emotions, whether directly, through communication, or through engagement with artistic depictions. These sorts of experiences refine empathetic capacity and thus work to correct empathy bias, and these experiences can be effortfully sought out. Therefore, empathy bias is not an innate and uncorrectable feature of our moral lives, but a problem that can be corrected with conscious effort to diversify experiences and develop wide-ranging and fine-grained emotion concepts.

Empathy bias can be characterized as a failure to empathetically connect with those whom we might learn something from, a failure based on arbitrary factors that have nothing to do with whether we should empathize. Insofar as I effortfully broaden my experiences such that I share them in some sense with others, then I try to make the factors contributing to the degree to which I can empathize less arbitrary. The goal of these three approaches is not to be able to empathize with as many people as possible, nor is this the goal of correcting empathy bias in general. The goal is to develop more nuanced emotion concepts such that one is better able to recognize potential avenues of connection with others, but this does not mean that one must be able to empathetically connect with *everyone*. Empathy bias is ultimately a bias regarding attunement to evidence: it prevents us from understanding the emotions of others that we can and should consider. There will always be disagreements that are not grounded in bias, but rather in

fundamental differences in values. Empathizing appropriately can help us recognize this in cases in which we empathize with someone to a degree yet still disagree with them. In terms of moral inquiry, the goal of the embedded, communicative, and imaginative approaches is to take steps to prevent biased empathetic blockages that prevent constructive moral evaluation. These blockages can be avoided by addressing the experiential blindness that underlies them. The three approaches to effortful correction of empathy bias discussed here enable one to find constructive avenues of empathetic connection based on shared experiences.

The aim of the effortful adjustment of one's empathetic capacity is to broaden one's experiences such that one has some level of conceptual overlap with the emotion concepts of those who may initially seem to be emotionally distant, and who will remain emotionally distant if one does not make the effort to broaden the experiential ground on which one's emotion concepts are based. Prior to such an effort one will not know how one's emotion concepts will or will not be adjusted, but insofar as one is aware of an experiential blindness, one can make the effort to correct that blindness via seeking out new experiences. This means that awareness of empathy bias is crucial to the correction of empathy bias. If one is aware of one's propensity to experience empathy bias, one can consciously correct empathetic blind spots by seeking experiences that might lead to more overlap with others' emotion concepts built up from similar experiences.

Importantly, this is not to say that the effortful correction of empathy bias means that one must achieve a complete conceptual overlap with another's emotion concept. One may have a similar experiential background to a target of empathy, but one will never have the *exact* experiential background of the target, and emotion concepts are nuanced enough that a slight lack of overlap may play a key role in the extent to which one can construct an emotion similar

enough to that of the target so as to find the other's perspective persuasive. It may be impossible to represent the emotion of another person with complete accuracy, given the practical infeasibility of taking into account that individual's entire unique conceptual schema and history, coupled with the unlikelihood that we will be able to entirely remove our own perspective from consideration. However, this does not mean that the correction of empathetic bias is impossible, as we must remember that empathy is a phenomenon that occurs in degrees. When I empathize with another person, a degree of overlap in emotion concepts may be all that is needed to lead to meaningful reevaluation of my own perspective, and that degree of overlap can serve as a catalyst for further efforts to understand the other. The fact that empathizing will not lead to a perfectly accurate representation of the other's perspective does not mean that we should not make the effort to represent that perspective as best we can, given that we do have the capacity to achieve a significant degree of conceptual overlap, and making the effort to achieve this degree of overlap can be greatly beneficial to moral inquiry in that we can achieve a degree of distance from our own biases.

Indeed, when it comes to the role of empathy in moral inquiry, there is an important middle ground between constructing an emotion that closely resembles that of the target, thus experiencing a high degree of empathy, and failing to construct any such emotion, thus failing to empathize. A capable empathizer could construct an emotion that resembles the target's moral emotion, but perhaps only to a weak degree such that the empathizer's own moral convictions are not in any way overturned. We can divide a meaningful effort to empathize with the moral emotions of another into three possible results: (1) The empathizer constructs an emotion that resembles the emotion of the target with a different moral outlook to a strong enough degree such that the empathizer significantly alters his or her own moral concept as a result. (2) The

empathizer constructs an emotion that resembles the emotion of the target with a different moral outlook, but not to a strong enough degree so as to alter the empathizer's own moral concept. Perhaps the empathy is not strong enough to alter one's moral emotions but is strong enough to motivate one to seek further experiential evidence to interrogate one's own moral view. (3) The would-be empathizer fails to construct an emotion that is anywhere near the moral emotions of the target despite an effort at nuanced perspective-taking based on relatively fine-grained emotion concepts, and this failure to empathize tells the would-be empathizer about the strength of his or her own moral concept. All three of these possibilities can be beneficial to moral inquiry if they result from a fallibilistic mindset and efforts to develop fine-grained emotion concepts grounded in diverse experiences. We need not universally aim for any one of these possibilities in advance of dealing with particular moral disagreements. The result of the effort to empathize will vary depending on the particular moral scenario in question and the character and views of the targets of empathy, and this is as it should be. The point is not that we should always aim for one of these results, but rather that we should aim at adhering to a *method* of nuanced, effortful perspective taking that draws on a wide array of experiences. It is this fallibilistic, experienced-based method of effortful engagement that accounts for the correction of empathy bias, not any sort of complete conceptual overlap with the emotions of another.

3.2.1: Does the Presence of Empathy Bias Preclude the Correction of Empathy Bias?

At this point a critic of empathy may be skeptical that one can break free of empathetic biases to begin the process of engaging in this method of refining emotions through experience in the first place. One might argue that, even if this sort of method could in theory correct empathy bias, the problem is that those who are already biased will not engage in the method. The concern is that empathy bias prevents the fallibilistic, experience-based method that might

correct empathy bias from ever getting off the ground. If we propose an embedded, communicative, or imaginative approach to correcting empathy bias, we need to be able to pursue these approaches in an unbiased manner. It will not do any good if these approaches do not extend past our ingroups; we will not be able to correct empathy bias if we only seek experiences, communication, and art that confirms our presuppositions and overlaps with familiar emotions, but this sort of self-confirming experience is precisely the sort of experience that biases may lead us to pursue. Thus, these approaches to correcting bias run the risk of establishing and perpetuating feedback loops in which we seek out engagement with experiences and values that are similar to our own and do not realize the benefits to moral inquiry that empathy affords us, the benefits of being able to reach outside of our own perspective and engage in moral evaluation from the standpoint of someone with a different perspective. This is a central problem facing an account of moral inquiry that emphasizes the benefits of empathy: how do we *motivate* the diverse experiential engagement needed to adjust empathy bias if the underlying empathy bias itself makes us less likely to pursue this sort of diverse experience?

Before outlining the embedded, communicative, and imaginative approaches, it is imperative to address this question. Again, because each of these approaches emphasizes the role of experience in shaping one's empathetic capacities, each approach runs the risk of shaping one's empathetic capacities in a biased manner if one does not pursue diverse experiences, that is, if one's choice of experiences is itself biased. The benefits of empathy for moral inquiry involve empathy's ability to enable one to evaluate one's own emotions, values, and morally relevant behavior from different perspectives. However, if we are not careful, we can cultivate an empathetic capacity that plays precisely the opposite role, empathy that encourages us only to evaluate our emotions, values, and behavior from the perspectives of those who already share our

presuppositions and limited experiential background. The aim of this chapter is to provide an account for how it is possible to correct empathy bias with effort. Such an account must answer this concern; it must show that the sort of effort required to correct empathy bias is not in fact precluded by the presence of empathy bias itself. CAT tells us that we can adjust our emotion concepts and by extension our empathetic capacity based on the experiences that we pursue, but are we capable of an unbiased pursuit of morally relevant experiences?

In 3.1 I argued that CAT provides us with the theoretical backdrop to understand how experience affects emotion and empathetic capacity. Insofar as we can choose our own experiences, then we can choose experiences that render us less biased by refining emotion concepts. So, if CAT is the right model of how emotions work, then, at least in theory, we can pursue the sorts of diverse experiences, whether embedded, communicative, or imaginative, that will leave us better prepared to empathetically engage with those who are not members of our in-group. However, in practice this can be difficult, as the very bias we need to correct can render us less likely to want to make this sort of effort.

What can motivate us to make this sort of effort? The key is awareness. One must be aware of three things: first, that one is in fact susceptible to empathetic biases; second, that these biases are grounded in emotions that can be adjusted via the pursuit of novel experience; and third, that these biases are harmful and worth correcting. Insofar as we do not want to be biased, understand that we are susceptible to (often implicit) empathy bias, and understand that our choice of experiences can do something to correct empathy bias, then we should be motivated to do the things that will help correct this bias, in this case pursue diverse experiences through the embedded, communicative, and imaginative approaches.

The psychologists Carol Dweck, Karina Schumann, and Jamil Zaki have explored the power of this sort of awareness to implement real change in empathetic capacity and behavior. Dweck's work focuses on how "mindsets," people's beliefs about their own psychology, affect their behaviors. Her research suggests that those who think of psychological characteristics such as extroversion or intelligence as fixed tend to attribute failures in this area to a lack of ability and will avoid opportunities for further training, whereas those who think of these characteristics as malleable skills that can be developed are more likely to pursue opportunities to improve.[64] Dweck, Schumann, and Zaki (2014) have applied this idea to empathy to ask the question: does having a mindset that it is possible to improve one's empathetic abilities result in greater efforts to empathize? Their results suggest that having a mindset according to which empathy is a skill that can be improved rather than a fixed trait results in behaviors such as spending more time listening to emotional stories from a member of a different race and devoting more energy to considering political opinions that differ from one's own. Their study also found that it was possible to change people's opinion on whether empathy is fixed or malleable. Participants presented with articles describing empathy as a skill tended to produce the behaviors associated with greater effort to empathize, while those presented with articles describing empathy as fixed tended to make less of an effort to empathize.

This research has significant implications for the correction of intergroup empathy bias. Drawing on Dweck's mindset theory, Goldenberg et al. (2018) found that Israelis and Palestinians who were encouraged to believe that groups are capable of change based on examples such as the Arab Spring and the formation of the European Union felt more positively towards the other side of the Israeli-Palestinian conflict and felt more hopeful about the possibility of peace, and that this change had a durable effect. This finding suggests that more

64 See Dweck (2006) for an overview of Dweck's work on mindsets. Also see Hong et al. (1999).

focused attention on the nuances of an outgroup's experience and history, coupled with an awareness that positive change is possible, enables one to better see a conflict from an outgroup member's point of view and to recognize shared goals.

Research on the role of mindsets in improving empathetic effort dovetails nicely with a CAT account of empathy. CAT provides a theoretical foundation according to which we can understand how empathy can be improved with effort, while research on mindsets suggests that an awareness of this possibility can have a tangible impact on how much effort we make to correct biases. Thus, the answer to the question of whether one's empathetic bias precludes the effortful correction of that bias seems to depend on how much one is aware of the existence of bias and of one's ability to correct bias. Empathy bias coupled with a mindset according to which empathy is not a skill that can be cultivated may hamper or preclude empathetic effort. However, an awareness that empathy is not fixed but rather is a malleable skill can enable one to pursue the sorts of effortful behaviors that will ultimately cultivate that skill. In this way, awareness of the prevalence of empathy bias and of the theoretical possibility that empathy can be cultivated by developing fine-grained emotion concepts can drive us to make the effort to correct empathetic biases.

We ultimately end up with a somewhat Aristotelean account of the effortful improvement of empathy, according to which seeing empathy as a virtue that can be cultivated will encourage us to pursue the sort of behaviors that do in fact improve our empathic abilities. As with any virtue, we ought to pursue a golden mean. We ought to try to correct empathy biases by seeking a greater range of empathetic engagement but avoid over-empathizing with too many perspectives such that we take on immoral views, and also avoid over-empathizing with too few perspectives such that we are closed off to the sort of critical perspective that drives further

moral inquiry and improvement. This Aristotelean approach to cultivating empathy, the possibility of which is supported by CAT and mindset theory, is crucial both in recognizing that it is within our power to improve upon our empathetic abilities and correct biases, and in calibrating how far we ought to go in trying to empathize with those who do not share our perspectives.

3.2.2: The Embedded Approach

The embedded approach to correcting empathy bias is a method of effortful, immersive experiential learning, but the end to be learned will not be understood prior to the act of immersion. While one ought to be conscious at some level that one's end is correcting empathy bias, the point of the embedded approach is to situate oneself in *novel* environments so as to correct for experiential blindness, meaning that one will not know how one's emotion concepts will be affected by such an environment prior to engagement. In terms of refining empathy for the purpose of moral inquiry, the novel environments that one experiences will be relevant to empathetic consideration of different perspectives on moral problems.

Two questions arise at this point. First, how do we identify which sorts of experiences are relevant to the moral perspectives of others; that is, how do we identify value-relevant experiential blind spots to pursue? Second, where do we draw the line regarding which sorts of value-relevant experiences we *should* pursue, given the fact that some values are not desirable targets of empathy?

The second question is no doubt extremely important, and applies not just to the embedded approach, but to the communicative and imaginative approaches as well. My response to this second question is largely the focus of Chapter 4, in which I defend the view that empathy

bias is worth correcting and that the correction of empathy bias ought to be motivated by fundamental considerations of compassion.

I do not think that we can answer the first question with a rule-based procedure by which we are able to identify a value-relevant experience in every case. It is often not clear to us which specific experiences have led us to arrive at our own values, let alone which experiences led others to arrive at their values. Values are complex beliefs built up from years of unique experiences and are in many cases not directly cultivated but rather are subconsciously influenced by culture and upbringing. Furthermore, emotion concepts are multifaceted, experience-based concepts that vary with individual experience, and the emotions tied to values, such as moral outrage, shame, and pride, are no different. What this means for the embedded approach to refining empathy is that we must settle for developing a general habit of cultivating an openness to new experiences, given that we may often struggle to precisely locate value-relevant experiences to seek out. However, while it may be difficult to locate which experiences are value-relevant for a given cultural or individual perspective, in cultivating a general habit of openness to new experiences, one remains open to experiences that may eventually play a role in empathetically simulating the emotions of those with whom one encounters a moral disagreement. There will be a greater likelihood of experiential overlap and thus a greater possibility of overlap of emotion concepts. As value disagreements arise, one who pursues the embedded approach will have a deeper well of evidence to draw on that might explain another's view and allow one to empathetically consider that view.

In this sense, the embedded approach involves consciously building up one's general empathetic skills, rather than consciously pursuing some specific perspective. One can proactively seek to build up a palette of fine-grained emotions through seeking out novel

experiences relevant to other cultural and individual perspectives, without necessarily having a particular value disagreement in mind. The idea is that exposure to such perspectives equips one to be able to empathize with more fine-grained emotions when moral problems involving individuals and cultures that one might otherwise be unfamiliar with arise. In terms of the theoretical backdrop of CAT, we can say that, in making the decision to immerse oneself in unfamiliar practices and environments, one subjects one's emotion concepts to potential experience-based alteration in the form of correcting prediction errors. This process refines emotion concepts to be better equipped to empathize when the need arises.

The most obvious way to engage in this proactive embedded approach is to travel widely, immersing oneself in different cultural practices and perspectives and sharing the experiences of a variety of individuals. There is perhaps no better way to understand another's values than to directly experience how the other lives, and this involves traveling to unfamiliar areas and experiencing different cultural practices and beliefs in action. As Mark Twain puts the point, "travel is fatal to prejudice, bigotry, and narrow-mindedness, and many of our people need it sorely on these accounts. Broad, wholesome, charitable views of men and things cannot be acquired by vegetating in one little corner of the earth all one's lifetime." However, an emphasis on travel perhaps makes the embedded approach the least practical of the three approaches to effortful correction of empathy bias. There is a sense in which the embedded approach asks one not just to imagine walking a mile in another's shoes, but to actually try walking a mile in another's shoes, with the goal that doing so will enable one to better construct the emotion of another when empathizing in the future. The problem is that it is often not feasible to engage in this practice at a widespread scale. Most of us do not have the luxury of constantly traveling around the world and directly experiencing a diversity of cultural and individual perspectives and

practices. As we will see in the following two sections, this difficulty can be effortfully remedied by the communicative and imaginative approaches. Regardless, I do not think it is a difficulty that rules out the feasibility and benefit of the embedded approach in general.

The reason for this is that, while widespread travel and engagement with diverse cultures and environments may often seem impractical, it is of course not impossible, and, importantly, the same sort of practice at a smaller scale is often not impractical and is still valuable. A less itinerant, more manageable proactive embedded approach could involve simply engaging with the experiences of segments of one's own community that one does not consider an ingroup. One does not necessarily have to travel widely to find novel and informative experiences. Valuable, emotionally informative experiences may stem from taking simple steps like attending a school-board meeting, volunteering at a local food bank, or attending any number of local community-organized events. While these sorts of small steps may not initially seem like much, emotion concepts may be built up from subtle experience over time. Rather than determine in advance which sorts of experiences are relevant or significant, the embedded approach is about developing a general mindset, a habit of seeking out experiences that are new and different, whether drastically or subtly so. This habit is not cultivated with some precise goal of achieving a specific emotional response, but rather is cultivated out of an awareness that one's ability to empathetically understand the emotions of others and remain open to the resulting moral and emotional growth relies on a general willingness to engage with novel experiences.

In terms of the feasibility of the embedded approach, it is also important to note that, oftentimes, our value disputes are not with those who are so vastly different from us, and they may not be the sorts of disputes that one would read about in a philosophy text. However, this does not make them unimportant or unworthy of careful reflection, reflection that can be greatly

aided by appreciating another's emotional perspective through a more fine-grained empathetic construction based on prior experience. Such reflection may be stymied by subtle empathy bias stemming from subtle experiential gaps. For example, you might have a dispute with the members of your local community over whether a new municipal building should be built in a certain location or whether a tax increase is justified. These are not momentous, philosophically significant value disagreements in the vein of disagreements over the ethics of abortion, or capital punishment, or restorative justice. Those sorts of disagreement are no doubt important and worthy of reflection as well, but smaller-scale value disagreements comprise a significant part of our everyday lives, and, importantly, they are the sort of disagreements that benefit greatly from making the effort to empathetically understand an individual's perspective as based in certain formative experiences. It is for this reason that a proactive embedded approach in which one builds on the potential for shared experience is so important. Building an experiential backdrop with more potential overlap with others enables one to empathize with perspectives that are worth consideration when trying to effectively solve these sorts of smaller-scale problems. Empathy bias exists at small scales as well. CAT tells us that the less experiential background we share with someone, the less equipped we will be to construct their emotions. Of course, there may be wide gaps between our experiences and the experiences of individuals whom we should empathize with, but there also may be smaller gaps, and we can take steps to close these smaller gaps as well with an embedded approach to pursuing novel experiences.

Critics of empathy often point to less subtle cases (e.g., deciding who to move up on an organ donor list, deciding to harvest one organ to save the lives of many, deciding how to distribute large amounts of charitable aid) when criticizing empathy's deficits. I agree that these cases can often highlight empathy's problems as a motivator for moral action; there are certainly

cases in which empathizing is not the best approach to solving a moral problem, and these extreme problems fit that description. But these less subtle cases are not the only moral problems that we face and in fact are not at all the sorts of moral problems most of us face in everyday life. Furthermore, as I argue in Chapter 4, empathy need not be the motivating force of morality to play a valuable role in moral inquiry. A point I want to stress throughout this dissertation is that there remains a significant role for empathy in subtler value disagreements, cases in which one can be drawn to critical self-reflection through empathizing with someone who holds a slightly but importantly different value in a concrete case in which action needs to be taken. The embedded approach, even when carried out at a smaller, more local scale, leaves us better equipped to consider other perspectives on problems that track that smaller scale.

Again, this sort of proactive embedded approach to refining our emotion concepts will necessarily be somewhat open-ended, as we will not know in advance what sort of perspectives and value disagreements we will encounter in the future, but the idea is that we should not be complacent in choosing our experiences in the present, because these experiences furnish us with useful emotional evidence that we can utilize when empathetically constructing the views of others. Though travel is helpful in this project, we do not necessarily need to travel to distant countries to implement an embedded approach; rather, what the CAT account of emotion helps us understand is that we merely need to act on the mindset that seeking novel experiences when it is prudent to do so, regardless of scale, can help develop emotional granularity and thus remedy empathy bias. Insofar as we value a lack of bias, this is a mindset worth adopting.

3.2.4: The Communicative Approach

The communicative approach to correcting empathy bias involves making an effort to directly communicate with those whose experiential backgrounds differ from one's own. The

goal, as with the embedded approach, is to subject one's emotion concepts to different experiences so as to refine the emotion and become better able to achieve some degree of conceptual overlap with those with whom one might empathize when engaging in moral inquiry. Unlike the embedded approach, the communicative approach does not involve experiencing novel situations directly; rather, it involves the process of imagining such situations based on directly engaging with those who have experienced them. In this way, the communicative approach lies somewhere between the embedded approach and the imaginative approach. Like the embedded approach, it involves some level of direct real-world experience, in that one seeks direct communication with others, but like the imaginative approach, it involves an imaginative projection of a novel experience rather than the direct experience of such a novel situation oneself.

The communicative approach can be particularly beneficial in appreciating perspectives on experiences that one cannot or should not pursue oneself, but that nonetheless are relevant to the critical examination of one's own values. For example, a white American cannot directly experience what it is like to be Black, a cisgender person cannot directly experience what it is like to be a transgender person, a non-immigrant cannot directly experience what it is like to be an immigrant, but of course these experiences are relevant to the moral, social, and political issues that one ought to consider. In addition to experiences that one cannot have, there are experiences that one should not directly pursue, but that are nevertheless important to understand when addressing moral problems, for example the experience of long-term homelessness or addiction. Given that one cannot or should not directly pursue certain value-relevant experiences that would improve one's empathetic capacity and reduce bias, one ought to find another approach to engaging with such experiences, and that is what the communicative approach

involves. It is important to note that we do not need to exclusively seek communication with those whose experiences we cannot directly access; there is value in communicating with those who have had experiences that we are capable of having or should pursue as well, insofar as that sort of communication can refine our emotion concepts and perhaps encourage us to pursue those sorts of experiences ourselves. However, the communicative approach is particularly well-suited to address the problem of building empathetic capacity based on experiences that one cannot or should not directly pursue.

It might initially seem naïve to think that merely communicating with those who come from different perspectives will make us better empathizers. What if we are too biased to seek out this kind of communication in the first place? And what if, even when we do so, our own emotion concepts are too deeply entrenched to be significantly altered by mere communication with someone who does not share our perspective? As noted in 3.2.1, the first question can be addressed by considering the awareness one can develop of the existence of bias and the mindset that one has towards improving one's ability to empathize. It is true that those who are biased, particularly those who are extremely biased such as white supremacists or anti-Semites, will resist the sort of communication that might improve their empathetic capacity and dissolve bias, but this does not mean it is impossible to do so, and there are ways to cultivate a mindset that values this sort of communication. Empirical work on mindsets suggests that if we can alter our mindset to consider our empathetic biases as correctable, then we will be better equipped to pursue that correction. Furthermore, empirical work in contact theory and conflict resolution suggests that communication plays a key role in initiating this shift in mindset and has the potential to result in a morally valuable improvement in empathetic capacity.

First introduced by the psychologist Gordon Allport (1954), contact theory is based on the simple idea that contact between individuals or groups in conflict can dissolve that conflict, and, importantly, can dissolve stereotypes and prejudices on which that conflict might be based. The idea is that contact with those who have different perspectives enables us to better understand the nuances of those perspectives, to understand "outsiders" as individuals with complex mental lives worth understanding and worth seeking common ground with, and to avoid the sort of sweeping generalizations that perpetuate bias. Since Allport introduced contact theory, it has been studied and adapted in a variety of ways, but the underlying idea that contact can reduce prejudice is strongly supported by empirical research. In a meta-analysis of 515 studies on intergroup contact theory, Thomas F. Pettigrew and Linda R. Tropp (2006) found that intergroup contact typically reduces intergroup prejudice across a broad range of group differences, including race, age, disability, ethnicity, and sexual orientation. Crucially, Pettigrew and Tropp (2008) found that empathetic perspective taking was one of three key factors (along with enhanced knowledge of outgroups and reduced anxiety in intergroup communication) involved in the process of reducing prejudice through intergroup contact, based on their meta-analysis.

So, there is evidence that direct communicative contact with members of outgroups is effective in reducing prejudice and in catalyzing empathetic responses. It is important to emphasize the communicative component of this contact. Contact with outsiders by itself, without any sort of constructive communicative engagement, will not have the same effect in improving empathetic skill, and in fact may merely amplify existing biases. If we do not take the time to constructively engage with members of perceived out-groups, but rather experience mere surface level contact that does not involve that group member's emotional perspective, we may

fall into the same prediction errors that allowed a misguided stereotype to develop in the first place. For example, Citrin and Sides (2008) find that non-immigrants in the United States and Europe tend to overestimate the number of immigrants in their countries and that this overestimation correlates with anti-immigrant attitudes, suggesting that more perceived superficial contact results in more antipathy towards immigrants. Similarly, Enos (2014) conducted a study in which Latino passengers were planted on a commuter train for 10 consecutive days. Enos found that white commuters who had been on the train grew less tolerant of immigrants than did passengers who did not ride with the Latino passengers.

These studies, and others like them[65] highlight the fact that contact alone does not lead to improved empathy. The context matters, and, importantly, the effort to communicate matters. Such an effort can correct misperceptions that, if left unchecked, are merely reinforced by surface level interactions. In a more neutral intergroup dynamic this effort should be easier to engage, but in cases of intergroup conflict, individuals might actively seek to avoid any effort to communicate and empathize with a perceived competitor or enemy, and this represents a significant obstacle to initializing a communicative approach to correcting empathy bias. For example, Porat et al. (2016) found that conservative Israelis tended to prefer not to empathize with Palestinians in general and this preference predicted lower levels of empathy when presented with concrete cases of Palestinian suffering.

It is difficult to conceive of a simple solution to such complicated intergroup dynamics, but this should not dissuade us from pursuing the benefits of a communicative approach to improving empathy and making progress where we can. The evidence in favor of the effectiveness of intergroup contact in reducing prejudice through empathy ought not be ignored merely because of the existence of some especially intractable intergroup conflicts. We ought not

65 See Hainmuller and Hopkins (2014) for an overview.

ignore, for example, findings that intergroup contact between Catholics and Protestants correlated with reduced dehumanization following sectarian violence in Norther Ireland,[66] or that white Americans who work or live with Blacks or Muslims are more empathetic towards Blacks or Muslims who are profiled by U.S. law enforcement,[67] or that white Americans with more empathetic contact with Black Americans are more likely to support the Black Lives Matter movement.[68] There is evidence of obstacles to improving empathy with intergroup contact, but there is also evidence that intergroup contact can be successful in reducing bias and driving solidarity. It would be a mistake to ignore this evidence of success and forego the potential benefits of a communicative approach to correcting empathy biases.

Furthermore, studying the ways in which a communicative approach can go wrong is helpful in that it will inform us how to approach intergroup contact to best improve empathetic response in each group. In studying how and when intergroup contact does or does not lead to improvement in empathy, we can understand how to create communicative contexts that facilitate the sort of empathetic development we want, and how to avoid communicative contexts that do not. Emile Bruneau's work is helpful in this regard. For example, Bruneau has found that intergroup power dynamics play a key role in the ways in which intergroup communication interacts with empathetic response. Bruneau and Saxe (2012) found that Mexican immigrants in the U.S felt worse about white participants after being asked to reflect and respond to essays in which white U.S. citizens wrote about "the difficulties of life in their society." The study also found that Palestinians felt worse about Israelis after being asked to respond to Israeli essays on their struggles. By contrast, when the roles were reversed and Mexican immigrants provided essays on their struggles, they felt better able to connect with the white Americans who read

66 See Tam et al. (2007).
67 See Johnston and Glasford (2017).
68 See Selvanathan et al. (2018).

them, and Palestinians who shared their essays felt better able to connect with the Israelis who read them. White Americans also felt better able to connect with the immigrants who shared their essays, and Israeli's also felt better able to connect with the Palestinians who shared essays. The takeaway is that empathetic connection between members of groups with differences in social privilege may be better facilitated by focusing on contexts in which the traditionally underprivileged group is given a greater platform to have their perspective heard. Feeling that one's perspective is genuinely heard breaks down an initial resistance to empathizing with the person who is doing the listening. In this way, the benefits of the communicative approach are not one-directional. Those doing the listening benefit from learning about the other's perspective, and those doing the sharing benefit from feeling better able to empathize in a communicative context that has been opened in a way that genuinely takes their perspective into account.

Understanding which approaches to intergroup communication are effective in improving empathetic engagement and which are not is an essential component of the awareness required to motivate active improvement of empathy. We cannot ignore obstacles that stand in the way of generating the sort of constructive contact with others that improves empathy, as these surely exist, but if we recognize that there are steps that we can take to address these obstacles by creating better communicative contexts, such as emphasizing shared goals and paying attention to power dynamics, then we ought to take those steps, given that communication does have the potential to improve our empathetic engagement. In communicating with others who do not share our perspective, we develop a better ability to empathetically simulate another's emotions based on their experiences, rather than merely projecting based on our own. We become more capable of other-oriented perspective taking and less focused on self-oriented perspective taking. This shift leaves us better able to incorporate valuable insights that we have not derived or

cannot derive from our own direct experiences as we try to determine the best approach to a given a moral problem.

3.2.4: The Imaginative Approach

While the imaginative approach to correcting empathy bias shares some of the features of the embedded and communicative approaches, it presents unique benefits for the correction of empathy bias and for moral inquiry. Like the embedded approach, the imaginative approach involves a certain immersion in novel experience. However, the experience in which one immerses oneself is an imaginative one, a projection into a fictional context. More in line with the communicative approach, imaginatively engaging with a fictional perspective involves engaging with experiences that one is not directly having oneself. Yet there are two features that distinguish the imaginative approach from the communicative approach, and these features are what account for the imaginative approach's unique benefits when it comes to moral inquiry. The first is the depth and pace provided by artistic representations of a moral perspective. The second is the creativity that artistic representation affords in depicting a moral perspective.

In reading fiction that explores the fine-grained emotions of its characters, we subject our own emotion concepts to feedback based on those characters' experiences. This sort of fiction encourages us to try on complicated emotional perspectives and supplies this empathetic effort with an in-depth consideration of the contexts in which the characters' emotions are situated. These are likely contexts that we have not directly experienced ourselves, but engagement with the fictional characters who experience them gives us a window into the sort of complex emotions that can arise in such scenarios. And literary accounts of these emotions allow us to explore them in great detail, as the mental landscape underlying even a single emotional experience can be expanded upon at length, a key emotional moment probed for several pages in

a way that the pace and etiquette of our interactions with others in real-time cannot always accommodate. We then incorporate this exploratory experience into our own emotion concepts and in turn develop a more wide-ranging, informed emotional understanding that is helpful in attempting to construct the emotions of other people when empathizing outside of fictional contexts. Because the psychological exploration involved in fiction is particularly deep, these experiences can refine our emotions in ways that merely communicating with others in our everyday lives may not. While, as discussed above, communication is helpful, we likely will not be able to appreciate another's inner-life in the way that we understand the inner-life of a character who is developed over hundreds of pages of a novel.

Furthermore, because this emotional exploration of character occurs in a fictional context, we can subject our emotion concepts to critical, empathetic examination by engaging with extreme perspectives or dangerous events that we might otherwise avoid. For example, when one reads Dostoevsky's *The Brothers Karamazov*, one can travel between Alyosha's vacillating religious ecstasies and doubts without joining a monastery, one can learn from Dmitri's manic moral exasperation without living out his Dionysian excesses or murderous rages, and one can explore the depth of Ivan's nihilistic angst without experiencing his psychic breakdown. Engaging with these characters as they experience moral dilemmas over the course of the novel allows us to explore these dilemmas ourselves, and to do so in a manner that refines our moral emotion concepts based on our experience of the dilemma through the filter of the characters' emotional lives.

In addition to facilitating this emotional depth and drawn-out pace of empathetic consideration, fiction has the capability to depict moral dilemmas in a creative manner that allows us to see them in a unique light. It is not just that fiction can plumb the depths of moral

dilemmas, but that it can do so in an artistic manner that even other forms of deep moral examination cannot.

Nussbaum (1983, 1985) argues that a particularly nuanced literary depiction of moral psychology is significant "not only in its causal relation to [a character's] subsequent speeches and acts, but as a moral achievement in its own right" (1985, p. 520). One of her examples is a scene from Henry James' novel *The Golden Bowl* in which the character Adam comes to recognize his daughter Maggie's autonomy and sexuality, realizing that despite his profound attachment to her, he must acknowledge her desire to start a new life with her husband, and must do so in a way such that Maggie does not feel guilt about leaving her relationship with her father behind. This is, as Nussbaum stresses, a "highly concrete" moral dilemma, rather than an abstract exploration of principles, and her point is that James' depiction of Adam's thoughts and actions surrounding the dilemma are a moral achievement insofar as Adam's mental life and the specific emotions behind his actions in this concrete case are a central aspect of what makes his handling of the dilemma morally commendable. The creative depiction of Adam's mental life enables us to understand the emotional conditions that play a role in his moral justification of the solution to the problem at hand. Nussbaum notes that James presents a lyrical explanation of Adam imagining Maggie as "a creature consciously floating and shining in a warm summer sea, some element of dazzling sapphire and silver, a creature cradled upon depths, buoyant among dangers, in which fear or folly, or sinking otherwise than in play, was impossible" (1985, p. 519). Nussbaum's point is that,

> If we had read, "He thought of her as an autonomous being," or "He acknowledged his daughter's mature sexuality," or even "He thought of his daughter as a sea creature dipping in the sea," we would miss the sense of lucidity, expressive feeling, and generous

> lyricism that so move us here. It is relevant that his image was not a flat thing, but a fine work of art; that it had all the detail, tone, and color that James captures in these words. It could not be captured in any paraphrase that was not itself a work of art. (1985, p. 521)

This is a key benefit of the imaginative approach to correcting empathy bias: there are certain important aspects of moral psychology that cannot be captured other than in artistic form. When it comes to improving our empathetic capacity, we ought to engage with art that eschews sententious rhetoric in favor of nuanced exploration of the emotional underpinnings of, as Nussbaum puts it, "moral anguish." Engaging with the layers of complexity involved in a character's experience and resolution (or lack thereof) of such moral anguish allows us to examine our own emotional response to the concrete moral dilemma depicted in a fictional account and to incorporate reflection on this fictional scenario into a more fine-grained emotion concept that is relevant to similar moral dilemmas. In the case of *The Golden Bowl*, perhaps our engagement with Adam and Maggie's moral dilemma in some way alters our emotion concepts involving things like parental love, or a desire for autonomy, and we can become better able to empathize with perspectives that value these emotions in ways that we had not previously examined. Fiction offers us a unique window into unfamiliar perspectives dealing with concrete moral issues, and we can use this window to understand emotions in a more nuanced manner, ultimately enabling us to appreciate that level of nuance as we empathize with different perspectives in our actual lives.

But at this point we should consider two important kinds of objections to this view of the value of empathetic engagement with fiction. The first kind of objection focuses on whether we really empathize with fictional characters in the first place. Carroll (2001, 2011) offers such an objection, arguing that our experience of engaging with fictional characters is not one of sharing

the affective experiences of the characters. The second kind of objection focuses not on whether we can, but rather on whether we should empathize with fictional characters. Serpell (2019), drawing on the work of Hannah Arendt, presents an argument that focusing on empathy as the locus of fiction's moral worth is misguided and that we should instead approach fiction with a mindset of "disinterested visiting." I will consider each of these objections in turn.

Carroll's criticism of empathetic engagement with fiction is that, "[w]e do not typically emote with respect to fictions by simulating a character's mental state; rather...we respond emotionally to fiction from the outside. Our point of view is that of an observer of a situation and not...that of the participant in the situation" (2001, pp. 311-312). Carroll offers several arguments in support of this view. The first is that our emotions cannot mirror that of the fictional characters because our emotions have different objects than do the emotions of the characters with whom we engage. The idea is that while a character might experience, for example, sadness over the loss of a loved one, we do not experience her particular sadness because it is not the lost loved one that is our object, but rather the character experiencing the loss. We feel sadness *for* her, but not *as* her. A second argument made by Carroll is that readers have access to additional information that affects their emotional experience, information that the fictional character does not possess. For example, we may feel fear for character X because we know that character Y is plotting to murder him, but we are not empathizing with X, who is in fact unaware of Y's intentions. Another related argument is that there can be an asymmetry between the desires and preferences of the reader and those of the character in question. For example, we might feel for the character, but nevertheless import our own ideas of how the character's moral dilemma ought to be solved and disagree with the character's preferred solution. Perhaps we understand a character's desire for his or her love interest, but we do not empathize with his or her jealousy or

anger when that love interest ends up choosing someone else; we may think that such a choice makes sense.

Responding to this objection with the conception of empathy outlined in Chapter 1 is helpful in illuminating both what I take empathy to involve and how it can be beneficial to moral inquiry when targeted at fictional characters. First, recall that a necessary condition of what I take to be empathy is self-other differentiation, and that empathy thus will always occur in degrees. When we empathize, we do not in fact take ourselves to be identical to the target of empathy, but rather experience some degree of the target's emotional experience, or some degree of some specific aspect of that experience in isolation of other aspects. So, while Carroll is right that there are often asymmetries between the emotional experience of the reader and the fictional characters in question, this does not rule out the possibility of important symmetries, and it is the process of challenging oneself to discover unexpected or particularly fine-grained symmetries that is at the heart of the value of the imaginative approach. One does not have to empathetically take on the character's entire identity, but rather may only take on significant aspects that ultimately end up challenging and refining one's moral emotion concepts. Self-other differentiation means that we can simulate a character's experience to some degree while maintaining our own emotional responses to that experience of simulation. Empathizing with a character who has experienced the loss of a loved one need not involve having that character's loved one as the object of our experience in the same way that the loved one is the object of the character's emotions, but it might involve experiencing some aspects of the emotional response to such a loss, and can still involve critically examining that experience through the filter of our own emotional background and life experiences. We need not take on the other's perspective entirely when we empathize, and in fact empathy's role in moral inquiry involves a balance of

taking on some aspects of the other's perspective while maintaining a critical distance from the experience so as not to be swept away by every empathetic experience such that we simply take on the moral views of every perspective with which we empathize.

But even if it turns out that we can in fact empathize to some degree with fictional characters, a critic might still challenge the value of such empathy. Does it really help us develop more fine-grained emotional concepts that make us better empathizers? Or is empathizing with fictional characters dangerous or ultimately ineffective in translating to our moral lives outside of fictional worlds? Serpell argues that empathy is merely,

> an emotional palliative that distracts us from real inequities, on the page and on screen, to say nothing of our actual lives. And it has imposed upon readers and viewers the idea that they can and ought to use art to inhabit others, especially the marginalized. (p. 5)

Serpell's claim is that our fascination with empathizing with fictional characters is a sort of counterproductive emotional tourism, problematic because it lures us into thinking we have achieved something morally because of the catharsis of empathizing with the suffering of others, particularly those who are marginalized, when in fact this catharsis does not translate into changes in the way we perceive or treat marginalized people in the real world. Drawing on Arendt's philosophy of fiction, Serpell argues that we ought to reject empathizing with fictional characters in favor of adopting what Arendt calls "representative thinking," a stance that considers the perspectives of others, not from within the other's perspective, but rather from a "general standpoint" that is achieved, as Serpell puts it, "by enlarging your mind to encompass the positions of others" (p. 7), rather than focusing on empathizing with one specific perspective. According to this view, we ought to approach fiction with the disinterestedness that comes from

this general standpoint in order to avoid the pitfalls of what Serpell takes to be selfish empathy tourism that does not translate into real moral worth. Serpell puts the point thusly:

> I find that the best way to grasp the distinction between "representative thinking" and emotional empathy is Arendt's lovely phrase, 'one trains one's imagination to go visiting.' This way of relating to others is not just tourism. Nor is it total occupation—there is no 'assimilation' of self and other. Rather, you make an active, imaginative effort to travel outside of your circumstances and to stay a while, where you're welcome. (p. 7)

Serpell goes on to emphasize that when one makes this imaginative effort to travel outside of one's circumstances, one nevertheless entirely maintains one's own identity; one does not empathetically absorb the perspective of the fictional character in question. But here again it is important to recognize that other-oriented perspective taking is not an all or nothing "assimilation" but a phenomenon that occurs in degrees, and that it occurs while maintaining self-other differentiation. In response to Serpell's point, we can ask: why must there be a strict binary according to which you either pursue Arendt's sort of disinterested visiting, or you pursue a complete assimilation of self and other? Keeping in mind what CAT tells us about the nature of emotions, both of these extremes may not be realistic, as one cannot simply turn off the emotion concepts that one has developed over one's lifetime, and one will not be able to achieve complete overlap with the unique emotion concepts of individuals who do not share one's exact experiences. But this is not a problem for empathy that is conceptualized in terms of degrees; it merely means that we ought to seek some balance of assimilation and maintaining our own distanced perspective. Indeed, targeting either one of these extremes seems more problematic than targeting this balance, as assuming you have reached either extreme can give you a false

sense of impartiality if in fact neither disinterestedness nor complete assimilation is realistically achievable.

The imaginative approach need not be in conflict with Serpell and Arendt's assertion that one should train one's imagination to "go visiting" in a fictional world. It is just that, according to the imaginative approach, the right kind of visiting requires a sort of humility regarding both our ability to achieve disinterestedness and our ability to achieve total empathy. Visiting a work of fiction requires one to make an effort to set aside aspects of one's own perspective, but not all aspects. Part of the value of engaging with fiction is that it involves an effort to achieve a critical distance while also empathetically engaging with the perspectives towards which one turns a critical eye. Pursuing some level of Arendt's and Serpell's disinterested "representative thinking" enables us to recognize the selfish, voyeuristic aspects of empathy that Serpell identifies, but there is still a pull to empathize that does not need to be fully ignored in the best works of narrative art; it is part of their aesthetic value, and it has value because of the sort of self-reflection on our own emotions that it can encourage. Serpell is right that "the idea that [artists] can and ought to construct creative vehicles for empathy… often makes for dull, pandering artworks" (p. 6). Yet, an artistic recognition of the tension between disinterestedness and empathy can lead to less didactic works that neither tell us that we ought to derive some ultimate objective disinterested moral truth, nor tell us that we ought to empathize completely with the moral views of its characters. Such works more realistically track the human experience of complicated, often inscrutable people that we nevertheless can empathize with to some degree and that we nevertheless must engage with when dealing with real moral problems. And fiction can shed light on these complicated characters in a unique and helpful way.

Furthermore, the imaginative approach is fully aligned with Serpell's calls for fiction to be more representative:

> Perhaps, instead of the current distribution—portrayals of "default humans" (that is, straight white men, good and evil) vs. empathy vehicles (that is, everybody else)—we could simply have greater variety of experience represented in our art. (p. 8)

The best fictional works for developing more fine-grained emotions will engender empathy without conceptualizing their characters as mere "empathy vehicles." Again, the key is valuing empathy in degrees. We benefit from some degree of empathy with a character in that it makes the work engaging and emotionally informative, but we also benefit from recognizing that there are aspects of any nuanced character that we cannot or should not empathize with,[69] and the coexistence of some degree of empathy with a character along with our critical, distanced consideration of other aspects is an important feature of the imaginative approach.

I have argued that the embedded, communicative, and imaginative approaches are viable routes towards the correction of empathy bias. With a mindset that takes into account our susceptibility to biases and a desire to correct these biases, we can utilize these experience-based approaches to develop more wide-ranging, fine-grained emotion concepts that enable us to better empathize when engaging in moral inquiry. The takeaway of this chapter is that empathy bias is in fact correctable with effort. CAT shows us that we can develop our emotional capacity by seeking novel experiences, and the embedded, communicative, and imaginative approaches suggest strategies for seeking these experiences. Nevertheless, as has been noted throughout, challenges remain regarding the potential dangers of this approach. Even if we can correct biases, is this process worth engaging in? Are the risks of empathy bias enough to make us

69 I explore this idea as it relates to "rough heroes" (often called antiheroes) in television programs in Kidder (2021).

neglect these avenues of potential correction and favor other moral approaches? In this chapter, I hope to have established that we can correct empathy bias, but whether we should make such an effort rather than pursue other avenues of moral inquiry is a separate question. Examining this question is the focus of the next chapter, and in the process of answering in the affirmative I will develop an argument for the unique value of empathy in moral inquiry.

Chapter 4

Responding to Empathy's Critics

The aim of this chapter is to directly address two of the strongest critics of empathy: Paul Bloom and Jesse Prinz. Bloom and Prinz present both empirical and conceptual critiques of empathy. Their critiques overlap in many ways, but they differ in terms of their favored alternatives to empathy. We can summarize their general argumentative strategy as follows: (1) establish that empathy is not a necessary feature of morality, (2) establish that empathy is problematically biased, and (3) conclude that we ought to pursue a moral approach that does not involve empathy. However, while Bloom advocates that we reject empathy in favor of what he calls "rational compassion," Prinz argues that we reject empathy in favor of other appropriately developed moral sentiments.

In discussing these views, I will, for the most part, agree with Bloom and Prinz regarding (1) and (2), but provide reasons to reject their inferences from (1) and (2) to (3). Without an effective critique of proposed alternatives to empathy-based morality, the overwhelming empirical evidence of empathy's susceptibility to problematic biases leaves the burden of proof on the defender of empathy's moral benefit. However, in finding good reason to reject Bloom and Prinz's critiques of empathy, I will defend the general value of the role of empathy in moral inquiry and show that the burden of proof that we ought not utilize empathy is shifted back to the critic and away from the defender of empathy's moral worth.

The structure of the chapter is as follows:

In 4.1, I consider Bloom's proposal of replacing empathy with rational compassion. I argue that there is in fact an important role for empathy within the general sort of approach to morality that Bloom advocates, and that role involves empathy's benefits for self-critical,

impartial moral inquiry. First, empathy enables us to critically assess our own assumptions about what the most compassionate, rational moral ends to pursue are. Deciding on rational ends to pursue requires not just an assessment of which consequences will result from a given decision or action, but also an assessment of the value of those consequences, and empathy enables us to step outside of our own perspective to consider alternative valuations and critique our own valuations. Second, empathetic engagement with others is in fact a compassionate response, as empathetically considering the moral views of others respects the individuality and authenticity of other human beings, their identity as moral agents of equal standing worthy of equal moral consideration. Furthermore, this compassionate response facilitates moral communication that opens up constructive avenues of critical moral inquiry.

In 4.2, I turn to Prinz's critique of empathy. In doing so, I do not take issue with his view that moral emotions play a prominent role in moral development, judgment, and conduct; this view is consistent with my account of moral inquiry. However, I argue that his critiques of empathy fail to take into account the role that empathy can and often does play in the development, critique, and calibration of the sorts of moral emotions that he favors as the foundations of morality.

In sum, my general strategy in this chapter will be to argue that, despite the evidence of bias considered in Chapter 2, Bloom and Prinz's rejection of a role for empathy in morality is too strong. The thread that runs through my critiques of Bloom and Prinz's views is that empathy is not only a means of detecting and responding to suffering in others; it is also a means of assessing one's own views and conduct from the perspective of another. While the former is usually the focus of discussions of empathy's role in morality, my claim is that the latter function of empathy, its role in self-critical moral inquiry, is its crucial function in the moral life. This

view is introduced in this chapter as a means of directly responding to Bloom and Prinz's critiques of empathy-based morality, and it will be expanded upon in the following two chapters as I outline the role of empathy in realizing impartiality defined in terms of fallibilist moral inquiry.

4.1: Empathy and Rational Compassion

Though Bloom does not quite explicitly tie his account of rational compassion solely to consequentialist ethics, his arguments against empathy presuppose that consequentialism offers the most impartial moral outlook, and he approvingly cites the sort of utilitarian reasoning favored by Peter Singer and effective altruists at a number of points.[70] Bloom's rational compassion clearly favors consequentialist reasoning, though he acknowledges that deontological principles do often play a role in our moral judgments and behaviors. In any case, the important point for Bloom is that *reason* can and should be the motivating factor in moral behavior, whether that reasoning involves the calculations of an act utilitarian or the rational application of moral principles.[71] Specifically, Bloom argues that reason should serve as a means of achieving compassionate ends, where compassion is defined as a general care and concern for the well-being of other people. Bloom puts the point thusly: "While sentiments such as compassion motivate us to care about certain ends—to value others and care about doing good—we should draw on [the] process of impartial reasoning when figuring out how to achieve those ends" (p. 51).

70 See especially pp. 102-106 and pp. 238-239.

71 Bloom does offer a brief defense (pp. 29-30) of the view that what may seem like Kantian consideration of moral principles can reduce to consequentialist reasoning. His argument is essentially a defense of basic rule consequentialism in which we ought to follow certain moral principles because in general, when these principles are followed there are better consequences. Bloom's work is not meant to adjudicate disputes about normative ethical theory, so we can set aside whether rational compassion is rational insofar as it follows principles or insofar as it appeals to consequentialist considerations. The important point for Bloom is that, in either case, it should be reason, and not empathetic engagement, that drives our moral judgment, development, and behavior.

Much of Bloom's book focuses on the sorts of problems and evidence discussed in Chapter 2 of this dissertation in an effort to prove that empathy presents an obstacle to what he takes to be the impartial reasoning that can render us ideal moral deliberators and actors. Though he recognizes that empathy is capable of motivating the compassion and altruism that he takes to be the fundamental goal of morality, Bloom's point is that empathy's "negatives outweigh its positives—and that there are better alternatives" (p. 241). The better alternatives consist in impartial reasoning motivated by compassionate ends such as "fundamental concerns about harm, equity, and kindness" (p. 239). Bloom argues that empathy is not necessary for morality because impartial rationality directed at compassionate ends can accomplish everything we want out of a moral system and can do so far more effectively then can an empathy-based approach. As Bloom notes, a key implication of this argumentative strategy is that his "antiempathy argument presupposes rationality" (p. 213), particularly in the moral realm. Thus, one way to critique Bloom's position is to critique the possibility of utilizing impartial reason in moral deliberation. Indeed, Bloom devotes the final chapter of his book to responding to just this sort of critique. However, this is not quite the strategy I will focus on here. Rather, I will grant Bloom the claim that we should utilize rationality in moral judgment but argue that empathetic, self-critical moral inquiry plays a key role in maintaining the impartiality Bloom wants to attribute solely to rationality. My claim is not that empathy is necessarily impartial; I hope that Chapter 2 has made it clear that this is not the case. Instead, I argue that reason is often not impartial either, as we have the potential to engage in motivated reasoning in our moral considerations. However, like empathy, reason is not irredeemably impartial. Thus, we are left with two potentially, but not necessarily, biased capacities vying for roles in our moral lives: empathy and reason. I argue that the solution to this problem is not to pit empathy and reason against each other such that the less

biased capacity ought to drive our moral lives while the more biased capacity is entirely neglected. Rather, my claim is that empathy and reason can work together to achieve a level of impartiality that neither could achieve in isolated operation. Thus, rational compassion need not exclude empathy and indeed is importantly bolstered by it.

In addition to arguing for empathy's role in attaining a sort of impartiality in moral reasoning, I will also argue that empathetic moral inquiry is itself the sort of compassionate end at which this reason is directed. It is hard to argue against the idea that our moral judgments and behavior should be motivated by a concern for others' well-being, that is, by compassion. The more interesting question is that of what we mean by the well-being of others: what does caring for another's well-being mean that we care about? I will argue that answering this question allows us to understand that empathy can play a significant role in considering the well-being of others, and thus empathetic engagement is a compassionate end. This is because of empathy's ability to allow us to focus on individual nuance rather than generalize about people as abstract entities. This may initially seem counterintuitive. If the goal is to develop a diffuse compassion for the well-being of humans *qua* humans, shouldn't we focus on what unites us as creatures worthy of ethical concern, rather than focus on the specific individual quirks that we often understand through empathy with particular individuals? In a sense, my answer to this is yes, but with the important caveat that a crucial part of what unites us as creatures worthy of ethical concern *is* our individuality, our capacity to care about unique values and projects and experience unique feelings. If this is true, then part of having a diffuse compassion for human beings involves caring about human beings' individuality and respecting their unique thoughts, feelings, and aims. Our ability to empathize allows us a particularly beneficial sort of access to these thoughts, feelings, and aims. To think that we can simply shut down empathy and reduce

ethics to consequentialist calculations based on compassionate values such as kindness, equity, and harm reduction is to fail to realize that empathy offers a unique form of kindness: it is kind to take the time to see things from another's perspective. It is to fail to realize that empathy offers a unique form of recognition of equity: to make an effort to inhabit the perspective of another is to recognize that perspective as importantly relevant. Lastly, it is to fail to recognize that empathy offers a unique means of reducing a particular kind of harm: to empathize with another is to make the other feel heard and meaningful in a way in which mere surface level acknowledgement of their views does not achieve, and in a way which can facilitate further constructive moral dialogue. In sum, once we understand the value of empathetic moral inquiry, empathy should not be opposed to compassion, but rather should be recognized as a compassionate concern for the individuality and moral agency of others.

4.1.1: Empathy, Impartiality, and Moral Reasoning

I want to argue that empathy is a means of realizing impartiality in moral inquiry in that it allows us to understand different perspectives on what the most rational, compassionate moral ends to pursue are. Of course we want to impartially arrive at the best moral decision, but we must be careful not to allow our own subjective tastes, values and life experiences to reify a standard of what counts as the best solution to a moral problem without allowing for other perspectives to weigh in on alternative solutions. This sort of consideration is not in the foreground when the moral judgments in question are such that most, if not all perspectives can agree upon what the ideal solution should be and what the ideal means of achieving such a solution are. But most moral dilemmas that we face in everyday life are not of this sort. My claim is that empathy contributes to impartiality when considering more subtle dilemmas. When there is disagreement about what counts as the most compassionate, rational moral decision,

empathy allows us to understand alternative perspectives, and to weigh our own view against the views of the individuals with whom we empathize. Empathy allows us to consider alternative value schemas as we assess potential solutions to a moral problem; in doing so, it allows us to see that these different value schemas may arrive at different conceptions of what the most rational approach to a given problem will be. Thus, empathy works with rationality insofar as it allows us to understand firsthand that our own view about the most rational moral solution is not the only possible view of a rational solution regarding the moral problem at hand.

In order to illustrate this point, we can consider the difference between two sorts of moral problems, one of which requires a fairly straightforward implementation of what Bloom would likely consider impartial reason, the other of which involves a variety of competing perspectives on what the most rational, compassionate solution would be. First, consider the decision to donate to a charity that purchases life-saving mosquito nets for those living in areas affected by malaria. Such a decision seems to be a clean fit with Bloom's criteria for impartial, rational compassion: it is compassionate in that it is motivated for a diffuse care for the well-being of others, and it is impartially rational in that it is motivated by consequentialist considerations of how one can do the most good. Importantly, this is an uncontroversial morally commendable action. The moral problem at hand involves what one should do to prevent the spread of malaria, and for one with limited financial means and medical expertise, donating a relatively small amount of money can save a large number of lives. Such a decision implements rational considerations regarding the impact of one's potential donation, and these rational considerations combat potential empathy biases that may lead one to spend the money on one's friends or family, or on causes more relevant to one's social ingroups, causes that would not have the amount of beneficial impact that donating to malaria prevention would have.

Now consider a second case. Suppose you are voting on your community's annual spending budget. The budget consists of spending in areas like education, environmental conservation, and supporting the development of local businesses. Each of these ends is compassionate in the sense that each does some good for some segment of the community, and it could even be argued that all members of the community benefit from any distribution of spending, just to varying degrees. But of course some segments will benefit more than others from certain distributions of spending. On the sort of effective altruist model that is the grounds for Bloom's rational compassion, the question we ought to ask is, which distribution does the most good? Which distribution is the most compassionate?

There are two ways disagreements may arise regarding this question. First, we might agree about what the ideal results of the spending would be but disagree about how distributing spending will accomplish this goal. Perhaps we have roughly the same valuation of the importance of education, environmental conservation, and economic development, but disagree about how much spending in each area is required to achieve our agreed upon goals. This is a practical disagreement about what the impact of spending will be. However, the second type of disagreement is a disagreement of values. That is, we may roughly agree about what the impact of a certain distribution of spending will be yet disagree about the value of that distribution. Put in terms of rational compassion, we disagree about which distribution is the most compassionate. While in the first case we may be able to follow Bloom's advice to apply reasoning to reach our shared goal of a particular end that we agree is the most compassionate, in the second case it is not clear how reason alone will tell us which end we should pursue. We agree about the facts about which ends that our decisions will lead to, and we agree that we want to direct our practical reasoning towards achieving the end that is most compassionate, but we disagree about

what that compassionate end actually should be because we disagree about the value of consequences.

This sort of problem does not only arise in these sorts of larger social dilemmas. For example, consider a case in which one is debating how best to care for an aging parent. The parent strongly values her independence and is resistant to moving to an assisted living facility, but her health is declining and there are significant risks involved in her continuing to live alone. In this case, the child and the parent might very well agree on the facts about what these risks are but disagree about how highly to value the experience of living alone. Suppose the aging parent is willing to accept the risk because she is very strongly committed to living independently, while the child places a higher value on safety and risk aversion. Here again is a case in which reason alone will not tell us what the appropriate solution is. Reason can only do so once we agree upon how to value certain considerations. If we assign a higher value to independence, then reason will tell us to pursue options that enable the parent to continue to live alone. If we assign a higher value to risk aversion, reason may tell us to pursue options that limit independence to a certain extent. The key is that this is a disagreement about values, and reason will not tell us how to value in such a case, but only how to realize values. As such, in order to pursue an impartial solution, we need to make the effort to try to empathize with alternative valuations that differ from our own, to recognize that ours is not the only possible rational solution, and to legitimately consider how we might see the problem differently from a perspective that values differently than we do. This is what empathetic effort enables us to do.

My goal in presenting these sorts of case is not to argue in favor of one particular solution. Rather, it is to highlight the sort of situation in which there are competing perspectives on what is the most rational, compassionate end to pursue, and it is not clear that we can argue in

favor of one perspective over the other without begging the question at hand. It is in these sorts of dilemmas that empathy can and should play an invaluable role, particularly if one's goal is to be impartial. Impartial reasoning in the moral realm should not involve paternalistically imposing one's own view of the most rational solution on all of the actors involved in the moral problem at hand. Rather, it should involve democratic communication about the value of alternative solutions, communication that involves legitimately considering multiple perspectives with a stake in different proposed solutions. While this sort of communication should certainly utilize rational consideration of the facts about what the consequences of each solution will be and how to realize certain ends, as I imagine Bloom would suggest, my point here is that there are many moral problems in which rational consideration of the facts about consequences alone will not yield agreement, because different perspectives on the problem are operating from different background assumptions about the value of certain consequences. Importantly, these background assumptions are influenced by one's culture and by one's unique individual experiences, including emotional experiences. As such, when seeking to resolve these sorts of disputes, the most impartial way of considering the other's view is not merely to consider how it measures up to one's own standards of what counts as a rational moral decision or what count as the best consequences, but rather involves making an effort to empathetically understand the other's different valuations, including their emotional underpinnings. In other words, an impartial approach involves making an effort to empathetically take on the perspectives of those whose purposed solutions differ from one's own when the difference in proposed solutions is not necessarily based in a disagreement about what the consequences will be, but rather is based on a disagreement about the value of different consequences. In making this empathetic effort, one makes an effort to evaluate one's own view in the light of perspectives that are different from

one's own and is thus making an effort to be impartial. One should try to empathetically understand the moral problem at hand from the other's perspective, taking into account their cultural and psychological background, in an effort to understand how they could arrive at a different view of what makes a particular moral solution the most rational solution, and in an effort to assess the merits of such a view.

While I am not disputing that the sort of irrational innumeracy that Bloom laments as a consequence of too much empathetic engagement poses an obstacle to unbiased moral judgment, numeracy alone, without any empathetic understanding of other perspectives, is not enough to overcome bias, and indeed can lead to biased rationalizations of one's own moral views. For example, a study by Kahan et al. (2017) found that those who rated higher in numeracy, a measure of the ability and disposition to make use of quantitative information, did substantially better than others in accurately assessing data related to a hypothetical clinical trial of a new skin rash treatment. However, the more numerate participants' assessment of the same data became *more* polarized and inaccurate than those who scored lower in numeracy when the data was linked to the partisan issue of gun control. In other words, those who rated higher in numeracy were *less* likely to accurately assess the data in a morally salient situation. The crucial takeaway of this study is that participants' rational numeracy skills were put in the service of maintaining their moral and political identities rather than in accurately assessing the data at hand. Kahan et al. call this phenomenon "motivated numeracy," and it poses significant challenges to the idea that some sort of unbiased, numerate rationality can impartially resolve moral problems. In order to impartially assess our own moral views, we must seek approaches that do not rely solely on the sort of numeracy-based consequentialist reasoning advocated by Bloom; failing to do so

leaves us at risk of merely rationalizing our own presuppositions about the best moral solution, rather than legitimately considering the possibility of alternative views.

Considered in terms of the general consequentialist message that one should seek to realize the most good, my point is that in the case of many moral problems it is not clear that there is one unassailable answer to what counts as the "most good," and empathy allows us to understand this in a way that reason alone cannot. Understanding alternative conceptions of the ideal solution to a given moral problem involves empathetically understanding the unique cultural and psychological factors at play in developing perspectives that value in ways that differ from our own ways of valuing. As a result of making the effort to inhabit these perspectives to some degree, one may radically or subtly shift one's own valuations regarding the issue at hand, or one may not revise one's own valuation. The important point is that in making the effort to see things from the other's point of view, a distinctive empathetic process, one is attempting to impartially evaluate one's own valuations from a perspective that does not share one's own biases or presuppositions. One makes the effort to evaluate one's own value schemas rather than merely evaluate the problem through one's own value schema. In order to impartially evaluate one's own conception of impartiality in the moral realm, to assess whether one is really being impartial without begging the question, there is a sense in which one should make an effort to get outside of one's own perspective, and this is precisely what empathy is uniquely suited to allow us to do.[72]

72 On my reading, this is one of Smith's central points in the *Theory of Moral Sentiments*. Smith's view is that we should try to empathetically place ourselves in the perspective of an "impartial spectator" when evaluating our own conduct. I discuss this idea at length in Chapter 5, arguing that Smith's view that an impartial spectator can be constructed by empathizing with others is problematic. I argue in Chapter 6 that impartiality is better conceptualized as a continuous fallibilistic method of inquiry aided by empathy, rather than a settled, idealized impartial perspective. Nevertheless, the key for both Smith's view and my own is that impartiality should involve empathetic engagement to assess our own biases.

I have been arguing that empathizing with different perspectives allows us to better understand alternative valuations in moral dilemmas and assess our own perspective with these alternative perspectives in mind, particularly in cases that are nuanced enough so as to not suggest near universal agreement on what moral solution is the most compassionate or rational to pursue. Bloom's critiques focus on showing that empathy is a problematic means of detecting and responding to suffering in others in an impartial manner; his arguments are meant to make the point that empathy often leads to harmful consequences. However, Bloom ignores the crucial role of empathy in impartially evaluating one's own view of what makes a particular consequence better than another, and this sort of consideration ought to be involved in moral inquiry, consequentialist or otherwise. We may seek impartial, compassionate moral ends, but in the process of establishing what those ends are, we need to be aware of alternative perspectives. We need to be able to imagine how our own particular cultural and psychological backgrounds may be influencing us to perceive certain solutions to moral problems as objectively rational or impartial, when in fact other moral deliberators from different backgrounds do not share our particular valuation of relevant consequences and thus do not share our particular conception of what ends are the most compassionate or rational to pursue.

While one can attempt to assess alternative perspectives without empathizing, one does so at the risk of imposing one's own perspective on the other. Absent any effort to empathize with other moral agents and deliberators, to *feel* the moral problem at hand from their perspectives, we will have a limited appreciation of the problem's nuance and of the appeal of other possible solutions from the perspective of others involved. Importantly, we will also have a narrow-minded justification of the impartiality of our own moral views. A truly impartial solution to a given moral problem must pass the test of assessment from outside of one' own

perspective. To empathize is to take on a perspective that is different from our own. Therefore, in order to impartially assess our own views, we ought to empathize with those who do not share our own psychological and cultural background so as to determine whether our views continue to hold the same weight in light of such alternative perspectives. Thus, truly impartial rational compassion ought to incorporate empathy, not necessarily as a means of detecting and responding to suffering, but rather as a means of assessing the impartiality of one's own moral assumptions from a perspective that may not share those assumptions. Even if we grant Bloom's consequentialist approach to moral reasoning, we see that empathy can help us understand alternative conceptions of what consequences are the most rational or compassionate to pursue. As such, empathy is an invaluable tool of moral self-assessment, even within the sort of consequentialist framework that Bloom favors.

It is also important to recognize that the fact that empathy occurs in degrees is central to its value in moral inquiry. We ought to assess our perspective and the problem at hand via empathizing with other perspectives, but this does not mean that we will necessarily be able to completely empathize with all moral perspectives, nor that we should be able to do so. The important point is that we make a legitimate *effort* to empathetically understand other perspectives, but this effort may very well leave our own initial valuations unchanged. There will be cases in which one may only be able to empathize with another's valuation to a small degree, or cases in which one may empathize with some aspects of another's perspective but ultimately be unable to empathize with the other's moral perspective at all. In any case, the effortful attempt to understand other perspectives subjects one's own values to critical test. One's values may pass this test or not, but if one is being impartial, one ought to make the effort to engage in the process of testing one's favored solutions in the light of other perspectives.

Empathy and reason ought to work together if we are truly seeking impartiality in our approach to moral problems. In Chapter 2 we saw that empathy, when left unchecked by rational considerations of broader consequences, is susceptible to crippling biases and innumeracy. Reason can help us overcome empathy bias, as in the case of donating to purchase mosquito nets. But we must also consider more subtle moral problems, cases in which agreement over what would be the most rational and compassionate solution is not so easily reached or intuitive to all those involved. These cases show us that if we do not interrogate *why* we take a certain moral solution to be the most rational and compassionate solution, if we fail to appreciate alternative valuations that are grounded in different experience and do not share our particular background assumptions, then there is an important sense in which we are not being impartial. What I have been arguing is that empathy can and should play a role in this interrogation of our own assumptions about what the most rational, compassionate solution to a given moral problem is.

I noted above that one argumentative approach to critiquing Bloom's account of rational compassion would be to highlight reason's own susceptibility to error and bias. As Bloom notes (pp. 221-225), there is an array of research in social psychology that purports to demonstrate the influence of unconscious, non-rational factors in our reasoning processes. In addition, there is our well-documented tendency to favor heuristics over more calculated, "Type 2" reasoning in a variety of applications.[73] While I have been drawing attention to reason's potential for bias in the moral realm, I agree with Bloom that social psychology research is not enough to suggest that we reject the role of reason in our moral lives, just as I do not think that the array of evidence of empathy bias is enough to make us reject the role of empathy in our moral lives. In his discussion of reason's susceptibility to error and bias, Bloom makes the point that, "even the most robust and impressive demonstrations of unconscious or irrational processes do not in the

73 See Kahneman and Tversky (1979).

slightest preclude the existence of conscious and rational processes. To think otherwise would be like concluding that because salt adds flavor to food, nothing else does" (p. 225). This is a legitimate point. However, I fail to see how Bloom has shown that this is not an equally legitimate point as applied to empathy and empathy bias. Even the most robust and impressive demonstrations of empathy biases do not preclude the existence of unbiased empathy. So, we are left with the following problem: reason and empathy are both deeply relevant to our moral lives and are both capable of facilitating impartial, morally commendable behaviors and views, but both reason and empathy have also been shown to be biased or error-prone in a variety of ways. The solution to this problem is not to reject the role of reason in our moral lives, nor is it to reject the role of empathy. The best way for empathy and reason to facilitate impartiality is for the two capacities to work together such that reason checks empathy bias, and empathy checks our assumptions about the nature of the most rational solution in a moral context.

My criticism of Bloom's case for rational compassion is not that it relies on reason, but rather that it pits reason and empathy against each other. In contrast to this approach, my claim is that reason and empathy should work together if our goal is impartiality. Reason can correct our empathy bias in a manner that directs our empathy in a more diffuse, impartial manner. Reason's role in recognizing the existence and prevalence of empathy bias is crucial in this respect. But empathy can reign in our tendency to assume that our particular solution to a moral problem is the only rational solution. Ideally balanced, this relationship between reason and empathy can serve as a virtuous cycle in which reason continues to expand our empathetic circle, and the expansion of our empathetic circle continuously and productively challenges our assumptions about whether we are being impartial when advocating for a particular solution as the most rational, compassionate solution when facing specific moral problems. Importantly, the

implementation of this virtuous cycle requires that we understand our biases and make an effort to correct them. In this sense the empirical work on these biases is invaluable. However, this work should be used as a motivation for developing a more balanced utilization of empathy and reason and should not motivate the complete removal of empathy from our moral lives.

At this point, perhaps Bloom may agree that rationality alone is not enough to ensure impartiality, but argue that it is not empathy that is needed to reign in reason's errors and biases in the moral realm; rather it is compassion that must play this role. One could argue that reason, when directed by our compassion for others, operates in the service of impartial morality. However, I think such a line of argument misses a crucial point that I have been stressing throughout this dissertation: empathy is not only a means of detecting and responding to suffering in others; it is also a means of taking on another's moral perspective so as to evaluate one's own views from outside of one's own biases. Furthermore, without empathetically assessing one's own view of the most compassionate solution to a problem from the perspective of others involved, one risks trying to check empathy bias via a biased conception of compassion. The key is that Bloom is focusing on empathy as a problem in terms of motivating moral action, whereas I want to stress that the appropriate role for empathy is not as a motivator for moral action, compassion can play this role, but as a tool for critically assessing the moral perspectives of oneself and others. The following section will expound on the significance of this distinction for moral inquiry.

4.1.2: Empathy and Compassion

In the previous section, I made the point that empathetically considering the moral perspective of others and assessing one's own moral views from another's perspective is an important part of seeking impartiality in moral inquiry. In this section, I want to make the claim

that this empathetic assessment is a compassionate moral response to others and thus ought to be pursued as part of a moral outlook motivated by compassion. So, I need not, and indeed do not, disagree with Bloom's emphasis on compassion in motivating moral action. Rather, my claim here is that it is compassionate to understand and respect others as unique individuals and moral agents whose views are relevant to moral inquiry, and that empathy enables us to experience and demonstrate this sort of understanding and respect. In this way, a desire to be compassionate should motivate us to be more empathetic. The empathy that I have been arguing we should pursue with moral deliberators involved in particular moral problems is not only a means of assessing our own impartiality; it is also a means of showing compassion towards those involved in the moral debate. Compassion towards others involves making an effort to understand things from their unique point of view, particularly when the unique points of view in question are the sort of significant, self-defining views that make up another's moral belief system.

Again, Bloom's idea of compassion is of a sort of diffuse care for the wellbeing of other humans *qua* humans. This is a fairly abstract concept, but I will refine it in a manner such that I take Bloom and other critics of empathy will find it uncontroversial. It is not my goal to define all the necessary and sufficient conditions for compassion, but rather is to recognize a feature of compassion that is aided by empathetic consideration of others' perspectives, particularly when those perspectives involve moral feelings and beliefs. The feature that I have in mind is that of concern for maintaining equity amongst human beings. In what follows, I will discuss empathy's role in realizing this feature of compassion, with the aim of showing that empathy should not be ignored even in the sort of compassion-motivated system that Bloom advocates.

I take it that maintaining equity amongst human beings is a key feature of compassion in that compassion involves care not just for those with whom one has close relationships, but also

involves a certain level of care for all human beings. In order to motivate this sort of diffuse care, one requires a deep respect for the equality of human beings. I will argue that a significant aspect of this sort of respect for human equality is a respect for each human being's capacity to formulate and carry out meaningful goals and projects, especially those related to moral issues. In other words, respect for human equality, and thus compassion conceived in terms of maintaining equity amongst human beings, requires respect for human nuance and individuality. If this is the case, then the role of empathetic moral inquiry as a compassionate response is clear: making an effort to empathetically take on another's perspective is an especially strong means of recognizing and respecting another's unique individual perspective.

What I have in mind here, then, is compassion motivating an empathetic appreciation of another's authenticity. In order to defend this view, I must first articulate what I take authenticity to mean. I define authenticity in terms of two claims that might initially seem to be at odds with one another, but which I think can be fruitfully united with empathy in mind. The two claims are as follows: (1) authenticity involves an agent's ability to determine his or her own life projects and commitments, and (2) authenticity is necessarily socially situated, rather than atomistic. In what follows I will argue that empathy plays a key role in respecting the authenticity of others, and thus that insofar as compassion involves respecting the authenticity of others, if we are motivated by compassion, then we ought to be motivated to engage in empathetic moral inquiry.

Let us first consider (1). A well-known proponent of this sort of view of authenticity is Bernard Williams. Williams' (1973) articulation of his view as part of his critique of utilitarianism is particularly relevant to my purposes here. His claim is that utilitarianism precludes the possibility of a moral agent's integrity, defined as an agent's ability to act with the motivation of furthering his or her own life projects and commitments. Utilitarianism precludes

integrity because within a utilitarian framework an agent's ability to act according to her own life goals and commitments is subjugated to a utilitarian calculation in which one must strive for optimific consequences defined in terms of the utilitarian conception of the good. If one's goals and projects are not consistent with achieving this optimific ratio, then one ought not pursue them. But Williams argues that integrity is a necessary feature of being a moral agent. Moral agency requires a particular point of view from which the moral problem at hand is understood and requires that a moral action be carried out because of one's particular moral views and character. However, because utilitarianism stipulates only *one* possible conception of the correct action, that which achieves an optimific ratio of pleasure to pain, utilitarian ethics does not involve individuals forming and evaluating unique sets of values based on their particular life experiences. For Williams, an essential part of moral agency is the ability to act in a moral situation such that one's action is motivated by one's own character, one's own deeply valued projects and commitments. If we remove this sort of motivation, then we remove integrity and thus neglect an essential feature of human authenticity.

It is important to note, though, what Williams is not claiming. His point is not that one cannot or should not arrive at a conclusion that a utilitarian calculus could endorse. Rather, his claim is that, if one is to arrive at such a conclusion and maintain integrity, then one must do so according to one's own personal commitments and values, and not according to a sort of blind adherence to the demands of a utilitarian calculus. Thus, to consider one of Williams' famous examples, if George decides to take a job developing weapons of mass destruction so as to best provide for his family, this can still be an authentic moral decision, and it is not impossible that it could be the right moral decision. However, it must be made according to George's own moral

commitments and character, and not because all actors ought to always act according to utilitarian calculations.

Framed in terms of compassion as discussed above, Williams' point is significant in that, if we care about the integrity of human beings, then we ought not try to impose a monolithic ethical theory on them; rather, we ought to enable them to pursue their own personal projects and moral commitments so that they can realize their authenticity as moral agents. This is where empathy plays a crucial role. For Williams, "the reason why utilitarianism cannot understand integrity is that it cannot coherently describe the relations between a man's projects and his actions" (p. 100). Man's projects are in an important sense irrelevant when assessing the morality of an action according to a utilitarian calculus; it is the action's consequence, rather than its relation to any projects, values, or commitments held by the actor, that is the relevant factor. By contrast, understanding this relation between one's projects and actions is precisely what empathy enables us to do. To make an effort to empathize with another's moral perspective is to take seriously her life goals and commitments. One may pay lip service to such goals, then appeal to universal principles in moral debate, but to do so would be to neglect the other as a moral agent, an individual with unique and relevant projects and goals. As I argued in the previous section, it is not enough to merely subject another's moral view to the evaluative criteria of one's own idea of the rational, compassionate solution, as those criteria may not be so impartial. If one appeals to one's own presupposed ethical framework to dictate what the appropriate moral sentiments for a moral actor to feel in a given situation are, we are still faced with Williams' question: "by what right does it legislate to the moral sentiments?" (1981, p. x). If one's integrity involves one's moral views and projects, then we can think of Williams'

question as asking: by what right does one individual's preferred value system violate another person's integrity?

In contrast to a violation of integrity, empathetically taking on the other's perspective validates the other's perspective as a legitimate potential contribution to the moral problem at hand. If we ignore the other's perspective in favor of an appeal to a static, codified ethical system, or if we only assess it in terms of how it relates to our own closely held values, then we are committing the same error that Williams identifies in utilitarianism: again, we are unable to "coherently describe the relationship between a man's project and his actions." Rather, we describe his actions in relation to the project of our own preferred ethical theory. On the other hand, if we empathetically consider the other's point of view, we can assess and respect the relationship between her actions and her projects, because we have made an effort to truly understand what those projects are from her point of view.

Again, I take it that compassion involves a diffuse care for others in the form of a desire to maintain equity amongst human beings. In defending Williams' view here, I have been trying to make the point that to hold others to a universal ethical standard without making the effort to empathetically consider the other's perspective is to ignore the other's moral agency, and thus is to fail to acknowledge their integrity in a significant sense. Thus, insofar as one values and respects others' identities as moral agents, as one should if one is being compassionate, then one should embrace a role for empathy in validating the significance of others' specific projects and aims; one should empathize so as to respect the other's identity as an authentic moral agent, to respect their perspective as a valuable input to moral inquiry.

At this point, the view I have been defending is subject to what I take to be a significant objection, namely that it encourages relativism and a respect for particularly abhorrent moral

views. I have been arguing that part of being compassionate towards another involves respecting the other's authenticity, and that this sort of respect is engendered by an effort to empathize with the particular goals and commitments that underlie the other's moral beliefs and behaviors. But what if those goals and commitments are, for example, racist, xenophobic, violent, or unabashedly selfish? Ought we still empathetically take on such perspectives in moral deliberation? If authenticity merely involves being able to formulate projects and to act according to those projects and commitments, does this mean that we ought to respect all projects and commitments if we are to compassionately respect authenticity? Does compassion dictate that we ought to empathize with those whom we take to be moral monsters?

These are serious questions that the view I have been defending has not yet addressed, however my hope is that that in defending claim (2) noted above, that authenticity is necessarily socially situated, I will be able to make some progress in addressing these concerns. In doing so, I will consider Charles Taylor's (1991) account of authenticity, in particular his defense of the role of "horizons of significance" (pp. 31-42) in realizing authenticity.

Taylor's goal is to defend the idea that one can reason with others about the nature of authenticity, that authenticity is not defined by the bare act of choosing *whatever* life projects and commitments one chooses. This point is essential in addressing the concerns regarding relativism discussed above. The key is what Taylor calls the "dialogical character" of human experience, that is, the idea that we define ourselves "always in dialogue with, sometimes in struggle against, the identities that our significant others want to recognize in us" (p. 33). Taylor is using "significant others" to mean those who "matter to us." However, in terms of our moral identities, the circle of who matters to us extends beyond family, friends, and partners and into the general population, as it is often those with whom we are not so close that we encounter

moral disagreement and deliberation; we often must interact and compromise with a diverse variety of individuals in order to solve a given moral problem and it is important to us that these individuals understand our moral views. In other words, we do not only want our moral beliefs to be recognized by those who are close to us, but rather want them to be recognized by those with whom we are engaged in practically relevant moral debate and morally motivated action.

With this in mind, we are in a position to understand Taylor's argument for horizons of significance and apply it to the moral realm. In doing so I hope to show that, while we ought to try to empathetically recognize the projects and commitments of others as part of compassionate moral inquiry, this does not mean that we must empathize with all individuals' moral views simply by virtue of the fact that those views are a part of an individual's projects and commitments.

Taylor's argument purports to show that appeals to authenticity that do not recognize "the demands of our ties with others" (p. 35), or that do not recognize the role of demands that extend beyond mere self-fulfilling human desire, are self-defeating. Taylor begins with the claim that defining oneself means identifying what is significant in one's difference from others. Making this sort of differentiation requires employing shared background assumptions about what sorts of things are significant. To use Taylor's examples, one cannot merely deem the fact that one has 3,732 hairs on one's head to be significant, or claim that the most significant, self-defining, and authentic action is wiggling one's toes in the mud. Some further story would be needed to make these sorts of assertions, e.g., that having that many hairs or wiggling one's toes connects one to some sacred spiritual experience. Taylor stresses that there is an implicitly understood distinction between the sorts of things that we take to be significant and those that we do not. This is what Taylor means by horizons of significance. In order for one to claim that we cannot reason with

one another about what counts as authenticity, one must claim that authenticity resides merely in the fact that one is choosing to identify something as significant. But Taylor's point is that to define authenticity as such is to collapse our horizons of significance. We want to be able to say that things such as one's creative talents or political beliefs are more significant than having a certain number of hairs on one's head or wiggling one's toes, but we can only do so if there is something beyond personal choice that is the criteria for significance. If we could merely choose what counts as significant, then creative talents, political beliefs, the number of hairs on one's head, and wiggling one's toes in the mud are on equal ground: if an individual deems them significant, we must respect them as significant. That is, if the mere fact that an individual deems something significant is enough to make it significant, then the very notion of significance is trivialized to the point at which we cannot make important distinctions between the sorts of things we want to say matter more than others. If significance is trivialized in such a manner, then the very notion of authenticity is trivialized, as authenticity is defined in terms of significant distinctions one can make between oneself and others.

The takeaway here is that defining authenticity merely in terms of one's ability to choose what one finds significant is self-defeating. As such, respecting one's authenticity does not mean that we need to respect one's view regardless of what that view is, merely by virtue of the fact that one chooses to hold that view and deems it to be significant. Rather, we can evaluate one's views in terms of horizons of significance. But this does not yet completely assuage concerns regarding moral relativism. Moral views, regardless of their content, seem to fall into the category of significant views, as they are clearly the sorts of self-defining views that Taylor identifies to be legitimate features of authenticity. So, the question remains: does respecting authenticity mean that we ought to try to empathize with any moral view on a given problem

simply because the view falls within the horizons of significance, regardless of how abhorrent we take that view to be?

In answering this question, I want to extend Taylor's analysis of horizons of significance to make the case for what I will call horizons of compassion. Roughly, the idea is that just as there are implicit socially agreed upon background assumptions regarding what counts as significant and what does not, there are socially agreed upon assumptions according to which we can judge a moral outlook to be compassionate or not.[74] Just as one cannot merely choose to call some attribute or view significant, one cannot merely choose to call any moral view compassionate; to do so would trivialize what we take compassion to mean. This is analogous to the manner in which authenticity is trivialized if it is defined merely by an individual's act of choosing, rather than by an individual's choices regarding questions of significance. If all that is required to deem a view compassionate is that the person who holds that moral view deems it so, then compassion no longer holds any moral weight. In calling one's own moral view compassionate, one's goal is explicitly to identify that view with helping others, but if a view is identified as compassionate *only* because one has stipulated that the view is compassionate, then consideration of others is not a necessary factor. One cannot simply stipulate that one's view is compassionate; something more is needed. The view must fit within horizons of compassion.

In determining which views meet such criteria, we can ask for *reasons* why a given view qualifies as compassionate. For example, we may ask how the view in question values the alleviation of suffering, how it respects equality, or how it exhibits kindness. If the view in question fails to meet any of our basic criteria for compassion, then we need not make an effort to empathize. My point is only that empathy should enter into our moral deliberation in situations

74 This is not the same thing as claiming that any view that is compassionate is morally correct or that it is the view that we should all hold. It is merely to claim that it has the feature of being compassionate. As we will see, one can recognize a particular view as compassionate but still prefer an alternative view.

in which different moral views both have some legitimate claims to compassionate solutions, but in which the solutions offered are different. As such, empathizing with others' moral views so as to compassionately respect their authenticity does not mean that one needs to indiscriminately empathize with all moral views regarding a given moral problem, but rather means that one should make an effort to empathize with those views that have a legitimate claim to be compassionate.

To clarify, it will be helpful to consider an example of how horizons of compassion can dictate which views we empathetically consider in a moral debate. Consider the issue of mass incarceration in the United States. As a result of harsh sentencing laws, the U.S. prison population grew by 222% between 1980 and 2010. Furthermore, the impact of this massive increase in incarceration disproportionately impacts people of color: while people of color make up 37% of the United States population, they make up 67% of the U.S. prison population. Black Americans are more likely than white Americans to be arrested, more likely to be convicted once arrested, and more likely to face longer sentences once convicted. These facts render mass incarceration a pressing moral problem that needs to be addressed, but there is not widespread agreement on what the best solution is.

Now, let us consider three perspectives on the issue of mass incarceration with horizons of compassion in mind. First, suppose that X argues that the mass incarceration of Black Americans is morally commendable. X has a racist moral outlook according to which Black people are inherently dangerous and should be subjected to longer prison terms and invasive policing strategies to keep the white population safe. When pushed, X declares that his view is compassionate, as it is motivated by his care for the safety of the white population.

Ought we make an effort to empathize with X's view in the process of evaluating our own moral view on mass incarceration? Does a motivation to be compassionate dictate that we respect X's authenticity by making the effort to empathize with his moral view? I think we ought not make such an effort, and the horizons of compassion enable us to justify this answer. Although X has stipulated that his view is compassionate, it clearly fails to meet the basic criteria of the horizons of compassion, as it ignores central considerations of equality and unjust suffering that are defining features of compassion. We can press X for *reasons* why we should consider his view to be compassionate. If these reasons are convincing, then we ought to make an effort to empathize with X's perspective so as to get a more fine-grained appreciation of his outlook on the issue at hand. In defending his view as compassionate because it is concerned with the suffering of whites, X is providing us with a potential reason to empathize. The problem for X's view is that his reasons are simply not good reasons to consider his view compassionate. The point here is that we need not and should not empathize with all perspectives involved in a moral dispute, but rather should empathize with those perspectives that can provide reasons to empathize in the form of legitimate reasons for considering their view to be compassionate based on the basic criteria stipulated by the horizons of compassion. In this way, efforts to empathize ought to be reason sensitive. If a moral agent's view is within the horizons of compassion, then one ought to make an effort to empathize with that moral agent, but if it does not meet that basic criterion, then one need not make such an effort. I have been arguing that a desire to be compassionate should motivate us to empathize with others involved in a moral debate so as to demonstrate respect for their authenticity by including their perspective as valued contributions to the process of moral inquiry. However, a desire to be compassionate should not motivate us to empathize with X, or with views that fall outside of the horizons of compassion, because such

views are clearly at odds with the ideal of compassion that is supposed to be motivating us to empathize in the first place.

Note, however, that just because we do not need to empathize with X's racist moral outlook when considering prison reform, this does not mean that we necessarily will fail to utilize empathy to recognize X's authenticity, his status as a human being with unique projects and aims. We may still make the effort to empathize with other aspects of X's experience in order to better understand his objectionable, uncompassionate moral view and its relation to his individual experience and psychology. For example, we might learn that X experienced a traumatic childhood of abuse and developed his views while seeking social refuge in white supremacist groups. We do not have to agree with X's racist moral outlook to engage in this sort of effort to better understand certain aspects of X's perspective through empathy, and making such an effort may be beneficial to moral inquiry in that it allows us to better understand the nuance of the moral psychology underlying such views, and to potentially recognize and open up avenues of communication that may change them. There is still a role for empathy to play in productive moral inquiry grounded in a respect for authenticity even when the target of empathy holds particularly objectionable moral views. However, that role is not to empathize with the uncompassionate moral views in question, but rather is to empathetically engage with other aspects of the perspective of the individual who holds that view in order to better understand how the uncompassionate view relates to that individual's identity and unique experience.

Now consider a different perspective. Y is deeply troubled by the implicit and explicit racism driving mass incarceration. She is concerned about the long-term impact on Black communities and families, arguing that the disproportionate incarceration of Black people for minor drug offenses exacerbates and perpetuates harmful racial inequalities in the United States.

In addition, Y has seen this harmful impact firsthand, as her father was imprisoned as a result of a minor drug offense and found it difficult to find consistent work upon release, leaving Y's family in constant worry about finances and access to healthcare. Suppose Y's view is that the United States should decriminalize drug use and pardon those individuals who are serving sentences for non-violent drug offenses. She argues that this will address racial disparities in policing and will drastically reduce the prison population. This reduction will have a significant financial benefit for the U.S. government and a significant social benefit for Black communities and individuals.

Ought we empathize with Y's view as part of the process of assessing our own view? Does a motivation to act compassionately mean that we should make an effort to empathize with Y's view so as to respect her identity as an authentic moral agent with a valued perspective relevant to moral inquiry? I think in this case one ought to make the effort to empathize and, again, we can appeal to the horizons of compassion to justify this response. Like X, Y claims that her view is compassionate, but unlike X, Y seems to be able to provide good reasons to support this claim. Her view is motivated by a desire to right racial inequalities and injustices and to help Black communities and families. It is a view grounded in the basic kindness and respect for equality that define compassion in general, and it does not include the discriminatory features that rule out X's view. Thus, if we want to respect Y's authenticity as a moral agent contributing to moral inquiry, then we ought to make an effort to empathize with her view if we are able to do so. Importantly, this means going beyond the sort of reasoning involved in determining whether or not Y's view is within the horizons of compassion. Once we understand a view as meeting this basic criterion, we can employ our ability to effortfully empathize using the embedded, communicative, and/or imaginative approaches to delve into the nuances of the view in question,

including the cultural and emotional backdrop upon which it is based. Seeking this level of nuance in the other's perspective is an important part of the process of demonstrating respect for the other's authenticity and opening valuable channels of communication. It is an important part of impartially evaluating our own views based on the idea that the other's perspective has the potential to make valuable contributions to moral inquiry based on their unique perspective and the valuations grounded in that perspective.

Now suppose that a third person, Z, is also opposed to mass incarceration, but that she argues that Y's approach goes too far. Z shares Y's concern about the racial injustice of mass incarceration, however her family has a history of drug addiction that she has seen ruin many lives. She worries that decriminalizing drugs will lead to higher addiction rates and more overdoses. As such, her view is that we ought to maintain the criminalization of drugs, but reduce sentencing for drug-related offenses and eliminate harsh provisions that increase sentences for multiple offenders.

Z's view is within the horizons of compassion as well. It is clearly motivated by a desire to alleviate suffering, as well as concern for equality and justice, and again it does not rely on the discriminatory premises that rule out X's view. Thus, we ought to make an effort to empathetically consider Z's view, both because doing so demonstrates respect for her authenticity as a moral agent with a valuable role in moral inquiry, and because her perspective can play a valuable role in our evaluation of our own view.

We can now see that while X's view does not fall within the horizons of compassion, both Y's and Z's views do fall within these horizons. Y and Z both have legitimate claims to compassionate moral views, but their views are different in important ways: they disagree on the value of decriminalization of drugs and thus favor slightly different approaches to addressing

mass incarceration via drug policy. While factual disagreements about the consequences of decriminalization may exist in this case as well, resolving the factual disagreement will not necessarily resolve the value disagreement between Y and Z, as the solutions favored by each are impacted by their different valuations of drug use and addiction. Y will be more willing to accept slight increases in drug use or addiction than Z.

Furthermore, their distinct views are deeply connected to their distinct cultural and individual psychological history. Empathizing with Y's specific emotional experience of frustration over her father's situation, or with Z's experience of grief over the loss of family members, helps us better understand these experiences as relevant in considering the moral problem of mass incarceration and drug addiction, and it also signals to Y and Z that we respect their unique personal experiences and perspectives and take them to be meaningful contributions to moral inquiry. Regardless of whether we ultimately favor Y's view, Z's view, or some other view that is within the horizons of compassion, the important point is that empathy allows us to do so based on sincere and compassionate efforts to understand alternative perspectives as authentic.

I have argued that making an effort to empathize with the moral perspectives of others is a means of recognizing their authenticity and value as a contributor to moral inquiry, and that this is a compassionate act that we ought to pursue if possible. Empathetically considering the other's moral views, rather than assessing them in terms of one's own moral views, is to take seriously what Williams called the relationship between another's particular projects and commitments and her actions; it is to respect her moral agency and open up valuable channels of communication and consideration that benefit an approach to moral inquiry that aims to be

impartial. Insofar as it is compassionate to recognize the equal moral agency of human beings, then we ought to try to empathize with those whose views differ from our own.

However, because authenticity is socially situated, because it is not defined merely in terms of what one chooses, but rather also in terms of socially defined horizons of significance and compassion, we need not seek to empathize with all moral perspectives. Rather, compassionate empathizing with other moral views should be directed at those who provide legitimate reasons for why their particular moral view is in fact compassionate.

In this section I have defended the value of moral inquiry based on two points that Bloom does not adequately factor into his case for rational compassion. The first is that if we want to strive for impartiality in moral reasoning, then we should not reason in a social vacuum, but rather should make an effort to empathize with other moral outlooks so as to assess our own conception the most rational, compassionate solution to a moral problem from the perspectives of others who do not share our particular psychological and cultural backgrounds and who may favor different valuations of the consequences in question. The second is that if we want to be compassionate, then we ought to empathize with the moral outlooks of those with whom we are engaged in moral deliberation, as this is a form of compassionate recognition of the authenticity and equal standing in moral inquiry that are central features of moral agency. To empathetically consider the moral views of others is to recognize others as legitimate moral agents with the potential to make valuable contributions to moral inquiry, and such recognition is compassionate insofar as it is grounded in care and respect for both the equity and individuality of human beings.

These two points are most salient in cases of subtle disagreement and in cases in which both parties involved are within the horizons of compassion, but I do not see this as problematic.

The moral problems we encounter in everyday life are for the most part dissimilar to the toy cases often employed by philosophers and psychologists; we do not face dilemmas over murdering innocent people to harvest their organs, or whether to push an obese man in front of a runaway trolley. Rather, moral debate often occurs regarding subtly, yet importantly distinct valuations that lead to subtly different preferred solutions to moral problems, as in the case of mass incarceration discussed above. My goal in this section has been to argue that we ought to keep an open mind regarding our own favored solutions in such cases, and that empathizing with those who disagree is a central aspect of maintaining such an open mind. Note that I am thus not defending the view that empathy is a necessary feature of all moral judgments. There are cases, such as the decision to contribute to the purchase of mosquito nets to prevent malaria, in which empathy does not seem to be playing any necessary role, even as a check on the rationality of such a decision. We do not need to empathize with perspectives that such a donation is irrational or immoral, as such perspectives fall outside of the horizons of compassion.[75] My claim here is only that these are not the only important sort of moral questions that we face, and when moral questions become subtler, empathy enables us to grasp different perspectives in a manner that is invaluable when seeking impartial solutions. If this is the case, then we ought not be against deploying empathy in our moral lives. We ought to try to use it effectively in the process of moral inquiry.

The thread that runs through my arguments in this section is that empathy is a tool that is best utilized in moral self-assessment; it is capable of allowing us to shed some of our individual

75 A possible scenario in which we should try to empathize with different perspectives is if another makes a case that our money would be better spent on some other altruistic cause. Such a perspective would fall within the horizons of compassion. This sort of debate would qualify as the subtler disagreement that I think requires empathetic self-assessment of one's own view; it is a very different sort of disagreement than that between one who argues that one should simply not donate one's money to charity at all and one who is considering donating to malaria prevention.

psychological and cultural biases when assessing our own views. As such, we ought to maintain an openness to empathizing with diverse views so as to leave our own views subject to potentially beneficial reassessment based on different perspectives that we encounter. While the empathy bias discussed in Chapter 2 certainly provides a barrier to this sort of open-mindedness, recognizing the unique benefits of empathy should motivate us to effortfully correct empathy bias, rather than to neglect the potentially fruitful role of empathy in moral inquiry. And as I argued in Chapter 3, it is possible for us to effectively make such an effort.

I turn now to Prinz's critique of empathy.

4.2: Prinz on Empathy and Moral Emotions

In this section I want to make two arguments in response to Prinz's claim that empathy ought to be avoided in the moral life in favor of other moral emotions. Before discussing my two objections to Prinz in more detail, it will be helpful to outline Prinz's general argument against empathy.

Although Prinz (2008) defends a sentimentalist view of morality in which emotions drive moral judgments, in his paper "Is Empathy Necessary for Morality?" (2011a) he rejects the thesis that empathy is in any way necessary for morality. He argues that other emotions such as guilt, anger, and compassion are the appropriate foundations of moral judgment, moral conduct, and moral development. His argument against a necessary role of empathy in moral judgment is based on providing examples of moral judgments to which empathy is irrelevant, for example the disapprobation of victimless crimes such as necrophilia or the desecration of the grave of a deceased person with no living relatives. It is also based on examples of intuitive moral judgments that seem to run counter to the judgment that he claims empathizing would suggest. Prinz discusses the often-used scenario in which one can harvest the organs of one innocent

person in order to save the lives of five others and argues that, because we should ostensibly have more cumulative empathy for the five suffering people than for the one individual who would be sacrificed, an empathetic moral judgment should favor the harvesting of the innocent individual's organs. However, our intuitive response is to disapprove of such a sacrifice; thus, we have a case in which our moral judgment runs counter to our empathetic response, according to Prinz. He concludes that, given that there are examples of moral judgments that do not involve empathy, empathy is not necessary for moral judgment. As noted in the previous section, I do not take issue with the claim that empathy is not necessary for moral judgment, nor do I take issue with Prinz's point that empathy is not necessary for moral conduct. However, I think these claims are relatively uninteresting in comparison to the question of whether empathy *ought* to be involved in *some* cases of moral judgment and conduct. Answering this question in the affirmative was largely the focus of the previous section.

Prinz's response to this question is to make the normative claim that empathy is "highly compatible" with social ills and "that should give us pause when reflecting on whether empathy is the key to a well functioning moral system" (2011a, p. 224). The social ills Prinz has in mind are the sort identified in Chapter 2 of this dissertation. These include "dangerous kinds of group thinking and intolerance" (p. 224), which he notes are often found in collectivist cultures that stress empathetic connectedness with other members, as well as the "dark side" (p. 224) of what Prinz takes to be the empathy-based morality of political liberals. He writes that "the politics of empathy tends to treat the victims of inequality without targeting the root causes" (p. 224), arguing that empathy-based social welfare policies ease the suffering of the poor, but are not effective in undoing the cycle of intergenerational poverty that has led the poor to require social welfare in the first place. The initial takeaway for Prinz is that "we should regard empathy with

caution, given empathetic biases, and recognize that it cannot serve the central motivational role in driving pro-social behavior" (p. 229). In a follow-up paper (2011b) Prinz's tone is less measured, as he writes that "empathy is, by and large, bad for morality" (2011b, p. 216).

At this point it is important to note that Prinz, like Bloom, is focusing on empathy's defects in terms of its ability to directly motivate pro-social behavior. I think Prinz is right to note the defects of empathy as the motivating force for all morality, but, like Bloom, in doing so he overlooks empathy's role in moral inquiry. Empathy's role ought to be one in which it is employed in the service of critical moral self-assessment of the principles and emotions that often do in fact directly motivate one's moral behavior. As I argued in the previous section, attempting to locate empathy as the motivating source of compassion reverses the appropriate direction of motivation: it is not that we ought to try to motivate compassion through empathy, but rather that our desire to be compassionate ought to motivate us to empathize with others in the process of moral inquiry. Furthermore, as argued in the previous section, an analogous point can be made regarding impartiality: it is not that our empathy motivates us to be impartial, but that our desire to be impartial should motivate us to empathize in a particular manner so as to check potentially biased presumptions. Thus, although empathy perhaps does not always play a direct motivating role in moral behavior, this does not diminish its value in maintaining levels of compassion and impartiality that are central features of the moral life; it merely means that empathy ought to function alongside other relevant capacities such as reason and other moral emotions. With this sort of role for empathy in moral inquiry in mind, we can now consider my two critiques of Prinz's approach to empathy.

4.2.1: Agent Empathy and Empathetic Self-Assessment

My first claim is that empathy can and should play a role in critically assessing one's own sentiments of approbation or disapprobation of certain morally salient actions and moral actors. This claim is a response to Prinz's (2011b) criticism of what he calls "agent empathy" accounts of moral judgment in which approbation and disapprobation are constituted by empathy (in the case of approbation) or lack thereof (in the case of disapprobation) with the motives of the agent performing an action. I do not support the strong thesis that Prinz is critiquing: the view that empathy plays a constitutive role in moral judgment such that to empathize with a moral agent is to approve of their moral sentiment and to not empathize is to disapprove.[76] However, I do think that making an effort to empathize with the moral sentiments of others is involved in our comparison of our sentiments toward a moral action and our sentiments towards the motivations underlying in it, and this comparative process is valuable to moral inquiry. As such, my issue with Prinz's critique is not that it rejects what he calls the "constitution thesis" about moral judgment, but rather that in ruling out a role for agent empathy in the phenomenology of moral judgment, Prinz characterizes agent empathy in a manner that seems to rule out any potential role for agent empathy in the phenomenology of moral self-assessment. Prinz ignores the role that agent empathy plays in our consideration of the relation of our sentiments of approval or disapproval of an action to our sentiments of approval or disapproval of the motivations of the agent performing that action.

We can see this by considering Prinz's basic example of approbation for another person who helps someone in need. He maps out the emotional geography of this approbation as follows:

> If I approve of your action, I will not feel gratitude. I will feel admiration. Gratitude and admiration are clearly different emotions. They have different

76 This view is held by Slote (2010a), who is the primary target of Prinz's critique of agent empathy.

> causes, phenomenology, and action tendencies. When grateful, there is a feeling of indebtedness and a tendency to reciprocate or express thanks. Admiration, on the other hand, has an upward directionality—we look up to those we admire—and tends toward expressions of respect rather than reciprocation. Admiration cannot be regarded as an empathetic response to the recipient of your generosity (the moral patient) because that patient feels gratitude and, perhaps, relief. Nor can admiration be regarded as an empathetic response to your motives (the moral agent). The feelings that motivate you are kindness or perhaps some anxiety or pity for the person in need. Admiration is not a feeling of kindness, anxiety, or pity. It is, again, a feeling with an upward direction. Pity is a feeling with a downward direction, and I certainly do not feel pity when I express approval for your act. My feeling is very different from yours. (2011b, p. 217)

I agree with Prinz that there is a sense in which the agent's feeling is different from that of one who approves of the agent's action. However, there are two important qualifications to add to Prinz's point: (1) the object of approbation in the case as described by Prinz is the action of helping someone in need rather than the underlying motivations of that action, but the psychology of approving another's action is not necessarily the same as that of approving another's motivations, including emotional motivations. I need not empathize with your pity when experiencing admiration for your action, but if the object of consideration is not the action itself, but rather the emotion that underlies the action, an effort to empathize can help me better understand the emotion (in this case pity) that I am evaluating. When we judge the moral emotions of others we must understand those emotions, and, as argued in Chapter 1, empathetic simulation appears to play a central role in this sort of understanding. (2) The fact that

admiration is not a feeling of kindness, anxiety, or pity does not mean that one cannot feel admiration *as well as* kindness, anxiety, or pity. Recall that the account of empathy I am defending involves affective matching and other-oriented perspective taking in degrees. Empathy does not involve a complete emotional match with another, but rather involves taking on some of the other's emotions to some degree while maintaining self-other differentiation. So, while the evaluator may experience an emotion of admiration that the agent does not experience, this does not rule out that the evaluator can also empathize with the agent's experience of pity or kindness towards the object of assistance. The evaluator may experience some degree of empathy with the agent's emotions, and this empathetic experience is part of the evaluator's particular instance of admiration for the action in question, adding to the admiration of the action because it provides a better understanding of the emotions that underlie it.

I think Prinz is right that insofar as the approval of the action is grounded in admiration for that the action and only that action, then that particular judgment does not require empathy with the agent. However, Prinz's example considers agent empathy in a manner in which its potential role is only in making moral judgments about the *actions* of the agent in question. If we take into account (1) and (2), we can recognize the value of agent empathy as a tool for making moral judgments about the emotions of the agent in question, judgments that are ultimately beneficial to the process of moral inquiry. This use of empathy as a tool to evaluate the relation between the moral emotions of others and one's own sentiments of approval in the process of moral inquiry is one that Prinz's critique of agent empathy, though effective against the constitution thesis, overlooks.

We can approach Prinz's example in terms of how empathizing in such a case could be involved in the process of considering the moral emotions of the agent, rather than merely

making a moral judgment about the agent's action. Constructing the phenomenology of this case with an eye towards judgment of the agent's emotions enables us to recognize the importance of agent empathy in a way that focusing only on judgment of the agent's action does not allow. Suppose I approve of you helping someone in need because I admire your action. If we stop here, as Prinz does in his consideration of the case, then indeed empathy need not play a role. However, suppose I press further and wish to examine *why* it is that I admire your action. This sort of second-order evaluative stance is of course not necessary to make the judgment of approval of the action, but it is the sort of stance that is necessary to assess one's own moral views in the process of moral inquiry. Notice that the object of moral consideration in the case of self-evaluation has shifted from your action to my approval of your action; I ask myself whether I should approve of my approval, rather than whether I should approve of your action.

Now, it might be argued at this point that although the object of evaluation has shifted to my own approval, this still need not involve any empathetic consideration of the agent's emotions, but rather may merely involve assessing the action at a deeper level. We might say that when evaluating the action, I ask, "do I approve of this action?", while in evaluating my approval of the action I ask something like, "what is it about this action that I approve of?" Perhaps the answer to this second question need not involve any understanding of the agent's emotions and thus not involve any sort of need to empathize. I could answer by saying that I approve of the action because it helped someone in need, which is something I value. In such a case I have interrogated my approval and engaged in moral inquiry but have not needed to empathize with the agent to do so.

It is true that we can ask questions about what relevant characteristics of the action warrant approval without empathizing with the agent, and this is valuable to moral inquiry in its

own way. But there is an important point to keep in mind here. Although there are relevant questions about our approval that involve only the characteristics of the action, this does not mean that there are not also relevant questions about the agent's motives, including emotional motives. The question "what is it about this action that I approve of," and "do I approve of the agent's motives underlying the action?" are distinct in that it is possible that I could approve of the action but not the motives underlying it, or *vice versa*. Nevertheless, the question of the agent's underlying emotional motives is still relevant to moral inquiry, if for no other reason than that it is important to understand that it may be possible for us to approve of a morally salient action while disapproving of the motivations underlying it, or *vice versa*. If it turns out that we approve of the action but not the motivations underlying it, this is an interesting fact about our moral psychology that is relevant to moral inquiry. We ought to ask why we approve of certain emotional motivations for certain actions and not others, or why we care about the relation between motivation and action in some cases and not others; doing so can productively challenge or reinforce our own moral sentiments of approval and disapproval.

An effort to empathetically simulate the emotions underlying a moral action can provide evidence that is relevant to my assessment of my approval or disapproval of the agent performing that action. If it turns out, for example, that the agent in Prinz's example actually experiences self-satisfied condescension or megalomania when helping another in need, I might be inclined to reevaluate my admiration of him. On the other hand, if my empathizing leads me to believe that the action was in fact motivated by kindness or pity, then I will approve of my initial admiration. In either case, the agent's emotional motivations are relevant evidence in terms of whether I ultimately approve of or revise my initial admiration, and, as I argued in Chapter 1, empathetic simulation enables us to gather this evidence, to better understand the

emotions of the agent in question.

So, Prinz is right that an empathetic experience of the agent's pity or kindness is not the source of approval of the agent's action; the source of approval in this case is admiration. However, if we subject that admiration to the process of moral inquiry by asking if we admire the underlying motives of the agent and not just his actions, then we should utilize empathy to gather relevant evidence. We can now see how this relates to (2) above. My admiration does not need to match with an agent's emotional experience in order to approve of his actions. Such admiration may coexist with a lack of empathetic consideration of the agent's motives, but it also may coexist with an empathetic experience of pity or kindness that is relevant, though not necessarily constitutive, of my approval of the agent's motives. Again, the key is to distinguish between the evaluator's moral judgment of approving of the agent's action, which requires only admiration and no agent empathy, and the evaluator's moral judgment of approving of the agent's motives, which is aided by the evaluator's empathetic examination of the agent's moral emotions. So, Prinz is right to reject the constitution thesis; agent empathy is not constitutive of moral judgment, nor is it even necessary. But stopping at this conclusion when considering the importance of agent empathy is a mistake, as there are moral judgments that are important in the process of moral inquiry, namely judgments we make about the relation of our own moral sentiments of approval or disapproval to the sentiments of the agents committing morally relevant actions, that do make use of empathy to gather relevant evidence.

4.2.2: Empathy Deficits, Shallow Affect, and Solipsistic Self-assessment

In this section I defend the view that self-critical, empathetic moral inquiry can and should play a role in the development of moral emotions that Prinz considers the foundation of morality. In considering the possibility that empathy plays a necessary role in moral

development, Prinz examines research on the emotional capacities of psychopaths. Psychopaths are distinguished by deficits in moral competence and also by their lack of empathy.[77] Thus, psychopathy offers potential evidence for the role of empathy in the development of moral competence. However, psychopaths are also distinguished by a general "shallow affect," a decreased capacity to experience emotion, especially the more complex sort of emotions involved in morality. Prinz (2011a, p. 217) approvingly cites Cleckley's (1976, p. 364) description of psychopaths' emotional capacities:

> Vexation, spite, quick and labile flashes of quasi-affection, peevish resentment, shallow moods of self-pity, puerile attitudes of vanity, and absurd and showy poses of indignation are all within his emotional scale and are freely sounded as the circumstances of life play upon him. But mature, wholehearted anger, true or consistent indignation, honest, solid grief, sustaining pride, deep joy, and genuine despair are reactions not likely to be found within this scale.

Prinz argues that empathy deficits are not necessary to explain psychopaths' deficits in moral competence, because "psychopaths will lack emotions that facilitate moral education as well as the emotions that constitute moral judgment" (2011a, p. 218). In other words, it is psychopaths' shallow affect rather than their lack of empathy that accounts for their moral deficits. In response to this claim, I will argue that Prinz is too quick to draw a hard line between shallow affect and empathy deficits. My claim is that empathy deficits lead to deficits in moral self-assessment in psychopaths. Psychopaths lack fine-grained, other-directed emotions precisely because they lack the empathy required to assess their own behavior from the emotional perspective of others. For example, the difference between the psychopath's "puerile attitudes of vanity" and the morally competent agent's "sustaining pride" is that the psychopath's vanity is grounded in solipsistic

77 Prinz cites Blair (1995) to support this claim.

concern for comparing himself to others, while the moral agent's pride is grounded in a genuine ability to empathize with how others actually view himself or herself. Thus, my point is not that Prinz is wrong to stress the significance of shallow affect in psychopaths, but rather that he is wrong to neglect the role that a lack of empathetic self-assessment plays in developing such a broader deficit in the moral emotions. Because of an empathy deficit, the psychopath is unable to reflect on his "shallow moods of self-pity" from outside of his own perspective so as to understand that they are in fact shallow and overly self-concerned. However, those who empathize with others' moral outlooks are able to engage in such reflection, and can thus refine their moral sentiments with broader society in mind. This is empathy's role in the "mirror of society" that Smith stresses as essential to moral self-assessment.

Prinz does not focus on the role of empathy deficits in limiting a psychopath's ability to understand the emotions of those who make moral judgments about his behavior and the behavior of others; rather, continuing the trend we have seen throughout both Prinz's and Bloom's critiques, Prinz focuses on how empathy deficits limit the psychopath's ability to understand the suffering of victims of moral infractions. He provides the following brief example of the sort of developmental story that is often "attractive" as an explanation for the psychopath's deficient moral sense:

> If a child with psychopathic tendencies hurts another child on the schoolyard and fails to experience empathetic distress, she may fail to understand why her behavior was bad. She might learn that teachers punish kids who harm others, but she will not understand what makes harm so bad in the first place (2011b, p. 221).

Prinz does not find this sort of story convincing. He offers the following response:

> I think this developmental story underestimates the resources that are available in moral

> education. Suppose a child is punished for hurting someone. The punishment may take several forms. She might be spanked, yelled at, sent to her room, or deprived of some privilege she enjoys. All these interventions will cause her to suffer. Aggressive punishment instills fear, deprivation instills sadness, and ostracism instills shame. In each case, she will also recognize that the love she depends on from her caregivers has been threatened, and the potential loss of love can be a source of considerable anguish. Moreover, children are inveterate imitators. A punished child will observe adult outrage at her actions and imitate that outrage when interacting with others in the future. In all these ways, the young transgressor learns to associate negative emotions with harm. But none of these forms of learning requires empathy. The victim often drops out of the picture as soon as the punishment begins. We might think of punishment as inculcating a sense of disapprobation directly without any essential empathetic involvement. (2011b, pp. 221-222)

With regard to psychopaths, Prinz's point is that moral deficits are more likely to be deficits in the fear, sadness, shame, anguish and outrage he lists as emotional responses to the punishments available as resources in moral education. His claim is that these particular emotion-based forms of moral learning do not require empathy because, even in the case of non-psychopaths, the "victim often drops out of the picture as soon as the punishment begins." So, a lack of empathetic consideration of the victim does not prevent the moral learning Prinz has in mind, yet psychopaths still do not experience this sort of learning. Therefore, it may initially seem that empathy deficits are not at the root of the psychopaths' lack of moral development.

Suppose we grant Prinz the point that empathetic consideration of the victim becomes irrelevant to the development of the moral emotions he has in mind as soon as the punishment

begins. I want to argue that this does not mean that empathy ceases to play a role in such learning. The key is that here again Prinz focuses only on the potential role of empathy for the suffering of a victim of an immoral act and neglects the role of empathy for the moral emotions of those who judge those who are responsible for such an action. Consider his point that ostracism brings shame. Why, we should ask, does the non-psychopath experience shame while the psychopath does not? True, this explanation need not involve the non-psychopath's empathy for the *victim* of the act for which she is being ostracized; but suppose the non-psychopath had *no* empathetic response in this situation. In what sense would we be able to say that she is experiencing shame? The experience of shame is one of understanding at a visceral level that others disapprove of your actions, that they are disappointed in you. We can provide an intuitive story about how empathy is involved in this experience. For example, suppose a child is ostracized by a group of friends after stealing another child's favorite toy. As Prinz rightly points out, the non-psychopath will experience shame based on being ostracized. But we must ask what makes this experience shame if not empathetic appreciation of the friends' moral disapprobation. True, the child need not empathize with the suffering of the child whose favorite toy was stolen, but it seems that he does need to have an emotional understanding of the disapprobation of his other friends, and this is what empathy enables.

By contrast, this empathetic understanding of others' moral emotions is what is lacking in the psychopath. The psychopath may feel self-pity or anger that his friends are ignoring him, and may even alter his behavior accordingly, but these are all self-concerned emotions that are not beneficial to developing a moral sense. His lack of empathy disables him from seeing this experience as a *moral* learning experience. Without an understanding of how others are experiencing moral sentiments regarding his behavior, the psychopath will not connect his

behavior to any sense of other-directed right or wrong, but rather will adjust so as to minimize detrimental personal consequences. The sting of empathetically understanding the emotional underpinnings of why others are ostracizing him will negatively affect the non-psychopath enough to discourage him from acting in similar immoral ways in the future, even if it is likely that he will not be discovered. His empathetic understanding of how others emotionally respond to such behaviors is a moral lesson that becomes a factor in his decision-making regarding similar actions in the future and his broader understanding of morally salient behavior. Empathy for the moral sentiments of others allows for an appreciation of the moral import of such actions, rather than a mere understanding of the personal consequences of being caught engaging in the behavior. On the other hand, the psychopath will assess the merits or demerits of such behavior in the future only in terms of whether he will face personal consequences (e.g., considerations of whether others will find out and deprive him of something he likes).

We can provide similar stories about the development of the other moral emotions Prinz discusses. For example, consider Prinz's point regarding the experience of anguish at the potential loss of love. Insofar as anguish is a moral emotion it should not be purely self-concerned. Anguish over a potential loss of love requires an emotional understanding of what the other's experience of love is like; empathetically understanding how powerful the other's experience of love is for oneself is what makes the potential loss of that love such a profound and morally educational experience of anguish. This experience aids in moral development in that it is a recognition of the connection between self and other on an emotional level; it can meaningfully impact one's views of the significance and value of the emotional lives of those beyond oneself. It is a way of recognizing that others do not merely interact with other people, but rather have deep emotional experiences towards them. Furthermore, it is a recognition that

oneself can be the object of such deep emotions as love. The psychopath, lacking empathy for the other's experience of love for himself or herself, cannot appreciate a connection between self and other at such a level. When the psychopath worries about a loss of love, it is not an anguished consideration of the emotions that the other feels towards him, but rather is a calculation of how the other will treat him if he is loved compared to how the other will treat him if he is not loved. This is not anguish but rather shallow self-concern and again, as in the case of shame and self-pity, the distinguishing factor between the two is a lack of empathetic consideration of the other's moral emotions as directed at oneself.

Lastly, consider Prinz's final point in the quotation above: that children are imitators and can learn to experience, for example, outrage by mimicking the outrage that they see others express. Presumably, the developmental story looks something like the following. A child harms another child and is then reprimanded by a parent or other adult. The adult expresses moral outrage that the child would engage in such behavior, perhaps by raising her voice or using certain emotionally charged phrases (e.g., "don't you ever do that again!"). The child then associates the act of harming another with this sort of chastising behavior but need not empathize with the victim of the harm to do so. So, when the child experiences others engaging in similar harming behaviors, she then mimics the sort of chastising behavior that the parent or other adult had directed at her.

The question to ask at this point, though, is this: how does mimicking behavior associated with outrage relate to the actual experience of outrage? If the child only mimics the chastising behavior when experiencing others impose similar harms but does not actually experience outrage, it is not clear that the child has incorporated a moral lesson into her behavior. Rather, she has learned how to behave appropriately according to certain accepted societal rules

regarding how one should respond to the harm in question. A psychopath can mimic the same sort of behavior but attach no moral significance to it. It can be socially beneficial to mimic this sort of outrage regardless of whether one actually holds a strong conviction or is actually experiencing any sort of strong moral emotion. One may do so simply to gain favor with a certain group or from a certain individual. But merely learning about the personal benefits of expressing outrage in certain harm situations is not a moral lesson.

By contrast, learning to experience an emotion of outrage in certain harm situations is a moral lesson. But it takes more than mere mimicry of displays of outrage to condition such an emotional response. Again, a psychopath can mimic outrage responses without actually experiencing the emotion. However, if empathy is involved in the process of associating outrage behaviors with certain harms, then one is able to tie the emotional experience of the other's outrage to the behaviors through which that outrage is expressed. If a child mimics a parent's displays of outrage because the child has empathized with the parent's outrage, rather than merely noticed the effects of its signals, then the child has incorporated a moral lesson, namely that the emotion of outrage is justified in certain cases of harm and can be expressed with particular signals. Importantly, empathy plays a central role in this sort of moral lesson. Empathy is the means by which the child associates the experience of the other-directed emotion of outrage with the behaviors by which it is expressed, rather than merely associating those behaviors with their ability to cause certain punishing behaviors in others.

My point in considering these examples is that moral development often involves the recognition of the impact of one's actions on the moral sentiments of those who make moral judgments, and not merely a response to the actions of those making the judgments. A self-directed appreciation that a certain kind of action leads to punishment, whether physical or

emotional, is not sufficient for moral development. A psychopath can develop in such a manner, avoiding those actions for which he thinks he will be punished either physically or by the experience of some shallow, self-concerned emotion. On the other hand, moral development involves the recognition that a certain kind of morally salient action is situated within a social context of individuals with complex moral emotions that respond in specific emotional ways to those actions. It is only in understanding that the people around you experience complex emotional responses of approval and disapproval of your actions that you begin to appreciate the moral import, and not merely the causal impact, of those actions. I have been arguing that empathy enables this understanding; it enables us to penetrate the surface layer of others' responses to our behavior and understand that we are often the objects of the moral emotions of others.

I must stress again that this learning process involves empathy for the moral emotions of evaluators, rather than empathy for the suffering of victims of morally deviant actions. The latter sort of empathy may indeed "drop out of the picture" in many instances of moral education, but the former ought not. Indeed, if one allows empathy for other moral agents' sentiments towards one's behavior to drop out, then one is at risk of developing the partial, self-serving tools that reach their extreme in the psychopath and making development instrumental and self-absorbed rather than other-directed and moral.

Prinz writes that through the resources of moral education he discusses, the "young transgressor learns to associate negative emotions with harm," but does not do so using empathy. I have been arguing that without the emotional appreciation of others' evaluations of one's behavior that empathy enables, what the young transgressor associates with harm are only certain punishing behaviors, and perhaps self-concerned negative emotions, not the sort of deep other-

directed emotions involved in moral development. Without empathy, the ostracized child associates the harm that caused him to be ostracized not with shame, but with a lack of play with others and with the shallow emotion of self-pity. Without empathy, the child who worries over a loss of love from a caregiver associates the harm that may lead to that loss with a lack of nourishing behavior from the caregiver and perhaps with the shallow emotion of selfish fear for his own wellbeing. Without empathy, the child who mimics her mother's outrage associates the harm that is the object of outrage not with the moral emotion of outrage, but rather with the ability to manipulate the behavior of others by displaying the physical signs of outrage.

The takeaway is that empathy's role in moral development is one of self-assessment, not based in taking on the perspective of those who are directly harmed or benefited by a certain action, but in taking on the perspective of those who judge our actions. Empathizing with those who evaluate one's behavior enables one to escape one's own solipsistic perspective on whether such behavior should or should not be pursued. A failure to escape this moral solipsism is the condition of the psychopath: it is an inability to critically assess one's own moral views and actions based on an inability to take into account the experience of empathizing with others' moral emotions as directed towards oneself.

In critiquing the views of Prinz and Bloom, my goal has been to advance an argument for the moral value of empathy based on the role of empathy in moral inquiry. While Prinz and Bloom are critical of empathy for the suffering of others as a motivational foundation for morality, they fail to recognize the value of empathy for the moral views of others in subjecting our own moral views and behavior to critical assessment. In 4.1 I argued that empathy can help us more impartially critique our assumptions about what count as the most rational, compassionate solutions to particular moral problems, and that it enables us to compassionately

recognize the legitimacy of the views of those who do not share our particular backgrounds as relevant contributors to moral inquiry. In 4.2 I argued that empathy enables us to evaluate why we experience sentiments of approval or disapproval for another's actions in relation to our sentiments of approval or disapproval of the other's underlying motivations, and that it enables us to develop deep moral emotions that are central to the moral life.

My aim has been to argue for benefits of empathy that Bloom and Prinz overlook in their wholesale rejection of its moral worth. However, it is important to note that I have not argued that empathy is not susceptible to the problems that Bloom and Prinz highlight. Empathy bias remains a pressing issue. As I have stressed throughout this dissertation, my approach is not to deny the moral significance of the evidence of empathy's susceptibility to problematic biases, but rather to defend a need to correct these biases. In critiquing Bloom and Prinz, my goal has been to articulate a role for empathy in morality that is important enough so as to warrant efforts to remedy empathy in response to evidence of bias, rather than to favor Bloom's and Prinz's favored solution of removing empathy from morality entirely.

This chapter has introduced such a role for empathy in moral inquiry in direct response to two anti-empathy accounts of morality: Bloom's rational compassion and Prinz's sentimentalism. I have argued that even these accounts should accommodate a role for empathizing with the perspectives of others in the process of moral inquiry. In the remaining two chapters, I will refine my account of this role, focusing on the relationship between empathy and the pursuit of impartiality.

Chapter 5

Adam Smith, Empathy, and Impartiality

In the previous chapter I defended the idea that a desire to be impartial ought to motivate us to empathize with a diverse variety of perspectives when exploring possible solutions to moral problems. One of my goals in that chapter was to outline a role for empathy as a helpful check on biases, a check that can facilitate beneficial moral self-evaluation by helping one achieve a more impartial perspective on one's own values. As such, I have been defending an approach to moral inquiry in which empathetic engagement is tied to a commitment to impartiality. Ultimately, empathy is beneficial to moral inquiry because of its in role within a method in which we reject complacency and remain critical of our own values; empathy enables us to remain open to new perspectives, and this fallibilistic approach is at the heart of what I take to be impartiality in moral inquiry.

In this chapter and in Chapter 6, I want to defend this role of empathy in a fallibilistic, impartial method of moral inquiry by considering two theoretical approaches to impartiality: Adam Smith's empathy-based sentimentalism and pragmatism. While Smith's approach presents some promising ideas, it also suffers from significant problems that need to be addressed in order to effectively tie empathy to impartiality. I will argue that these problems can and should be addressed by an approach to empathetic moral inquiry that is influenced by pragmatism, especially the work of John Dewey and Jane Addams. Thus, the goal of this chapter and Chapter 6 is to defend an approach to empathy-based moral inquiry that incorporates some insights of Smith's empathy-based account, but ultimately locates impartiality in a commitment to fallibilistic method rather than in the realization of the perspective of the sort of ideal "impartial spectator" that Smith describes.

Again, my claim is that commitment to an impartial, fallibilistic method of moral inquiry should motivate us to empathize. This is in an important sense an inversion of the way that empathy is often thought about with regard to motivating moral action. Defenders of empathy often posit the following sort of role for empathy in morality. When we empathetically experience the suffering of someone who is harmed, we are able to take that suffering into account in our moral judgment that the harm is wrong, even though we are not the objects of the harm. Empathy enables us to recognize the significance of the harm and motivates us to impartially arrive at moral judgments based on an action's effects on others and not just on ourselves. Furthermore, because the experience of empathetic suffering is more visceral than merely contemplating the suffering of another, empathy may be more likely to encourage moral action.

I defend a different sort of role for empathy, which can be roughly described as follows. When we make moral judgments about the rightness or wrongness of actions or views (including our own actions or views), we ought to try to empathize with the views of others who are involved in the moral problem at hand: those who are in a position to make a moral judgment about the action or view in question (again, including our own actions and views).[78] The goal of empathizing in moral inquiry is to enable us to see the action or view in question from a variety of perspectives and weigh the merits of those perspectives against one another, rather than assess the action or view only from our own limited perspective.

78 Note that this will include the views of those who are directly affected by the action, but that my claim is that one ought to empathize with these individuals' *moral views*, not necessarily with their particular experience of suffering, pleasure etc. While their views will almost certainly be informed by their suffering, pleasure etc., they will also be informed by other experiences and background views, and it is important to focus on trying to empathize with this more well-rounded evaluative perspective rather than merely with the suffering, pleasure etc. caused by a particular action.

So, I take the goal of impartiality as a starting point and argue for the conclusion that if we want to be impartial then we ought to make an effort to empathize with the moral perspectives of others. This may initially seem problematic given the evidence of empathetic bias discussed in Chapter 2. However, while that evidence tells us that empathizing to achieve impartiality faces obstacles, it does not tell us that it is impossible, as we saw in Chapter 3. My goal in this chapter and in Chapter 6, expanding on the considerations of the previous chapter, is to provide the philosophical grounding needed to show that this effortful empathy is worth pursuing as a means of achieving impartiality; if we want to be impartial, then we ought to incorporate empathy in a fallibilistic approach to moral inquiry. In order to defend this claim I will situate my view in relation to Smith's empathy-based account of impartiality, and to the pragmatist approach, which considers ethics as concerning the application, evaluation, and potential revision of values in response to, in Dewey's terminology, "problematical situations."

In this chapter, I will discuss Smith's account of the empathetic construction of the "impartial spectator" that he defends as the ideal moral perspective. I provide an interpretation of Smith's sentimentalist ethics according to which this impartial perspective is grounded in an empirical method of empathizing with and amalgamating a diverse variety of perspectives, rather than in realizing an abstract *a priori* ideal. While I think Smith is right to argue for the empirical, social construction of impartiality, I highlight a number of problems that Smith's theory encounters. Specifically, Smith's account of the impartial spectator is open to concerns regarding bias and relativism. His particular account of empirical, social, empathetic moral inquiry is subject to a pressing question that we have seen arise throughout this dissertation: *which* perspectives ought we empathetically consider when pursuing impartiality? This question is especially important given what I have outlined in Chapter 2 regarding evidence of empathy

bias, as such bias leaves us more likely to ignore certain perspectives. I outline a number of difficulties for Smith's account that are based on this question and argue that Smith's particular approach to empathy-based impartiality is at a loss to respond to these difficulties. My goal in making this argument is to highlight the difficulties that an empathy-based account faces if it tries to establish a static ideal of impartiality based on empathetic engagement, rather than motivate an empathetic approach based on commitment to fallibilistic inquiry.

My discussion of Smith in this chapter leads to Chapter 6's discussion of pragmatist ethics, which I take to offer an approach that can accommodate the benefits of Smith's empathy-based method of empirically and socially constructing impartiality, as well as address concerns about Smith's method's susceptibility to bias and relativism. My aim is not to mount a thorough defense of pragmatist ethics, but rather is to argue that commitment to the pragmatist values of fallibilism, democracy, anti-absolutism, and empiricism favor an ethics that is both empathy-based and defines impartiality in terms of method. Commitment to these values requires the sort of proactive empathetic engagement that Smith discusses as a central feature of moral inquiry. While this may initially seem to leave pragmatism vulnerable to the same concerns about empathy bias and relativism that appear to plague Smith's sentimentalism, I argue that utilizing a pragmatist-influenced approach leaves one equipped to provide a response to these concerns: a response that does not jettison empathy from morality, but rather seeks to utilize awareness of bias to appropriately correct empathetic engagement when addressing particular moral problems. I argue that a rejection of rule-based approaches to morality, as well as an emphasis on the role of individuality in the evolution of the moral self, provide a response to objections that plague Smith's empathy-based method of moral inquiry. In answering these objections with a pragmatist-influenced account, my goal is to defend the role of empathy as central to achieving

impartiality as defined in terms of fallibilistic method.

The structure of this chapter is as follows. In section 1 I summarize Smith's account of the impartial spectator, focusing on its empirical and social construction and on its relationship to moral self-assessment.

In section 2 I present objections to Smith's account of impartiality that are specifically based on the empirical and social construction of the impartial spectator. My goal in doing so is to highlight potential problems facing the sort of empathy-based account of impartiality that Smith provides. The problem for Smith's view is that defending the development of impartiality as an empirical, empathetic process leaves open questions regarding which perspectives should be considered in constructing an impartial spectator, and attempting to answer these questions based on Smith's proposed moral framework in which we ultimately arrive at a fixed, ideal impartial perspective pushes one towards relativism and accentuates issues regarding empathy's susceptibility to bias. I highlight two problems facing Smith's empathy-based system, problems that must be addressed when defending any empathy-based account of moral inquiry: (1) Such inquiry may lead one to construct a biased conception of impartiality, as empathy bias leads one to only consider a certain kind of perspective as relevant in the empathetic construction of the impartial spectator. (2) Such inquiry may lead to moral relativism, as it is not clear how the method itself is able to decide which of two competing perspectives is the more impartial without begging the question or succumbing to an infinite regress.

I address these concerns in Chapter 6 by defending a pragmatist-influenced method of empathy-based moral inquiry that accommodates the insights of Smith's account regarding the empirical, social construction of impartiality, but is also suited to address concerns regarding bias and relativism, particularly due to its shift away from general moral rules and towards

addressing particular moral problem situations through commitment to a fallibilistic method, and its insistence on recognizing the significance of individuality in solving moral problems and striving for moral growth.

In addressing the objections to Smith's empathy-based account while defending a role for empathy in fallibilistic moral inquiry, my aim in this chapter and Chapter 6, building on the arguments of the previous chapter, is to argue that empathy plays a valuable role in impartial moral inquiry.

5.1: Smith's Account of Impartiality

5.1.1: Self-Evaluation and the Impartial Spectator

In *The Theory of Moral Sentiments*,[79] Smith outlines a sentimentalist explanation of moral evaluation according to which appropriate moral judgments are carried out by adopting the sentiments of an ideally impartial witness of the conduct in question. He argues that it is only when we adopt the perspective of the impartial spectator, a perspective that is stripped of what Smith calls our "self-love," that we can properly make important distinctions between blame and blame-worthiness, and between praise and praise-worthiness. In other words, for Smith, it is through the perspective of the impartial spectator that we determine what is "the natural and proper object of praise" (p. 114), rather than what is merely the object of our own praise. As such, the impartial spectator plays a central role in Smith's sentimentalist framework. Smith argues that we ideally adopt the perspective of the impartial spectator both when we evaluate the conduct of other people, and when we evaluate our own conduct (pp. 109-110).

In this section, I will focus specifically on the role of the impartial spectator in Smith's account of moral self-assessment. According to Smith, if we perform self-assessment appropriately, then it is not a potentially biased self that does the evaluating, but rather is an

79 Hereafter referred to as TMS.

impartial spectator that has been constructed via empathetically considering the moral sentiments of others. While the impartial spectator may be the “man in the breast,” it is not constructed by internal self-reflection, but rather is dependent on empathetic interactions with others. Thus, proper moral self-assessment for Smith is really a sort of inter-subjective evaluation in which each individual evaluates moral actions and opinions based on the shared moral sentiments of others, which are understood by empathetically taking on their perspectives. This is Smith’s means of outlining a universal standard of impartial moral judgment within a sentimentalist framework; proper moral judgment, including self-assessment, is that which is conducted through taking the perspective of an impartial spectator, and this perspective is constructed through empathetically considering the moral perspectives of others.

For Smith, the construction of the impartial spectator is an empirical, social process that involves amalgamating the perspectives that one encounters in the world, the perspectives with which one empathizes. Smith writes: "We must become the impartial spectators of our own character and conduct. We must endeavor to view them *from the eyes of other people, or as other people are likely to view them*" (p. 114, my emphasis). In order to "become" the impartial spectator that ought to perform moral self-evaluation, we have to evaluate ourselves "from the eyes of other people," and to do so requires an empathetic understanding of the actual sentiments of the people in our society.

I think there is much to be said for Smith’s view regarding the need for empathizing in developing impartiality. In particular, as I argued in the previous chapter, I think that Smith is correct that impartial self-assessment must be a social, empathetic process rather than a matter of mere internal self-reflection and potentially biased confirmation of one’s moral assumptions. However, my goal in discussing Smith’s account here is in part to draw out how such an account

of impartiality remains subject to significant issues regarding bias and relativism, issues that need to be addressed if one is to advocate an empathy-based account of impartiality. Before discussing these issues, it will be necessary to outline the relevant details of Smith's account that leave it open to such objections.

5.1.2: The Empathetic Construction of the Impartial Spectator

It is clear from the outset of Smith's discussion of self-assessment in Part III of TMS that society has a large role to play. After briefly stating the need for impartiality in self-evaluation in 3.1.1 and 3.1.2, Smith shifts to a discussion of the role of society in constructing such an impartial view. He begins 3.1.3 with the passage with which I began this dissertation:

> Were it possible that a human creature could grow up to manhood in some solitary place, without any communication with his own species, he could no more think of his own character, of the propriety or demerit of his own sentiments and conduct, of the beauty or deformity of his own mind, than of the beauty or deformity of his own face. All these are objects which he cannot easily see, which naturally he does not look at, and with regard to which he is provided with no mirror which can present them to his view. Bring him into society, and he is immediately provided with the mirror which he wanted before. (p. 110)

Smith's point here is that society is the mirror with which we can recognize the moral nature of our own conduct. Our situation within a society of other moral actors allows us to exercise our own moral sentiments of approval or disapproval regarding the conduct and dispositions of other people. Crucially, when we engage in such moral judgments, we realize that those who we judge are judging us in the same manner.[80] It is important to note how this process

80 This is the sort of realization that I argued, in the previous chapter, is lacking in psychopaths.

relates to Smith's claim that each of us naturally desires a harmony of sentiments with the sentiments of other people—what Smith calls "the pleasure of mutual sympathy" (pp. 13-16)—and his claim that we approve of the sentiments of others only insofar as we can sympathize with them (pp. 16-19).[81] Early in TMS, Smith writes: "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" (p. 16). As social beings, we recognize *ourselves* as objects of the approval and disapproval of other human beings, and to evaluate whether the sentiments of others are "suitable" evaluations of our conduct, we must try to take on the perspectives of the people who are judging us so as to see if we can empathize with their moral sentiments towards us. Smith's point is that once we recognize that we are in fact being judged by those around us, we should seek to evaluate our own conduct and views as if we *are* those other individuals (p. 112), and should incorporate the insights gained from this empathetic process into our self-assessment. In empathizing with the sentiments of others, we can bring our sentiments regarding our own conduct into alignment with the sentiments of other people—we can attain the "pleasure of mutual sympathy." Our embedding within a society allows us to see our conduct through the eyes of others, and this empathetic perspective is the "mirror" through which our conduct is most clearly presented for "suitable" evaluation on Smith's view.

But there is still a question of what exactly the relation is between seeing our conduct with "the eyes of other people" and seeing our conduct from the perspective of the impartial spectator. After all, the perspectives of those who are judging our conduct may not be impartial. In considering this question, it is crucial to emphasize that the process of self-evaluation described above is an *empirical* process. As Smith notes, a human being who matures "without

81 This claim is analogous to the constitutive approach to agent empathy that was discussed in the previous chapter, and that both Prinz and I reject. Again, Smith's use of the term 'sympathy' can be thought of as analogous to what I mean by 'empathy.'

communication" with other people would be entirely incapable of self-evaluation. If self-evaluation is impossible outside of society, then impartial self-evaluation is certainly impossible outside of society. On this account, we do not possess an inherent concept of impartiality, but rather must fashion one, and our tools for doing so are the actual sentiments of those around us. Smith writes that "the eyes of other people" are the "*only* looking-glass" (p. 112, my emphasis) with which we can "scrutinize the propriety of our own conduct" (p. 112). The point here is that the impartial spectator must be derived empirically; it is a concept that is learned from our experience within society as we evaluate others and empathetically assess others' evaluations of our own conduct and views. The perspective of the impartial spectator is not an abstract concept that is distinct from seeing "with the eyes of other people,"[82] but rather is empirically constructed based on the moral sentiments of people that we actually encounter.

It is in this way that Smith argues that virtue and vice "have an immediate reference to the sentiments of others" (p. 113). On Smith's view, we must judge our own character or conduct as virtuous or vicious from the perspective of an impartial spectator, and "[v]irtue is not said to be amiable, or to be meritorious, because it is the object of its own love, or of its own gratitude; but because it excites those sentiments *in other men*" (p. 113, my emphasis). In other words, in order to properly evaluate whether one's own character or conduct is virtuous, one must construct an impartial spectator based on others' sentiments of approval or disapproval; one must

82 It is worth noting here the frequency with which Smith uses "the eyes of other people" and similar phrases interchangeably with and in close proximity to the term "impartial spectator." Consider, for example:

> Whatever judgment we can form concerning them, accordingly, must always bear some secret reference, either to what are, or to what, upon a certain condition, would be, or to what, we imagine ought to be *the judgment of others*. We endeavor to examine our conduct as we imagine any other fair and *impartial spectator* would examine it. (p. 110, my emphasis).

> But, in order to attain this satisfaction, we must become the *impartial spectators* of our own character and conduct. We must endeavour to view them *with the eyes of other people*, or as they are likely to view them. (p. 114, my emphasis).

understand that one's conduct or character is the object of genuine approval in others if it is said to be truly virtuous, and this understanding is achieved through empathetically considering oneself from the perspective of others.

Smith's point here is that virtue and vice are concepts that can only be defined in terms of how character and conduct relate to other people:

> To be amiable and to be meritorious; that is, to deserve love and to deserve rewards, are the greatest characters of virtue; and to be odious and punishable, of vice. But all these characters have an immediate reference to the sentiments of others… The consciousness that [virtue] is the object of such favourable regards, is the source of that inward tranquility and self-satisfaction with which it is naturally attended. (p. 113)

When one engages in self-assessment of one's moral character and conduct, one only reaches the point of "tranquility and self-satisfaction" when one comes to understand that other people genuinely approve of one's character or conduct. There is an important distinction to be made here. It is a distinction that Smith is careful to emphasize throughout his account of the impartial spectator: that between being approved or disapproved by another and being *deserving* of another's approval or disapproval. If one is to engage in impartial self-assessment, then one ought to consider not just whether one is praised or approved by others, but rather whether others' have good reason for such approval.

Smith's point is that, in determining whether one's own actions are truly virtuous or vicious, one should make an effort to empathize with the sentiments of those who approve or disapprove of the actions in question so as to discern the underlying motivation for such approval or disapproval; this is one's means of separating deserving approval or disapproval from mere approval or disapproval. If it turns out, for example, that those who approve of one's conduct do

so because they hold some prejudiced belief or emotional response, or only because they think it will personally benefit them in some way, then one ought to revise any feelings of tranquility and self-satisfaction that may have initially accompanied recognition of the other's approval.

For example, suppose that X is critical of a politician. Y expresses approval of X's critiques. Initially X may feel buoyed by such approval—she may feel validation that her opinion is virtuous as a result of its approval by other people. But to cease reflection at this surface level is to miss the point that Smith has in mind with the impartial spectator. In order to impartially evaluate her feeling of self-satisfaction, X ought to make an effort to empathize with Y and with those who disagree with Y so as to discern whether Y's approval was justified and impartial. Suppose that in making such an effort, X comes to realize that Y's approval was in fact motivated by a racist prejudice against the politician in question, whereas those who disapproved of X's critiques did so on the basis of fact-based policy matters. In making an effort to empathize with Y, X now finds it difficult to affectively match with Y's approval, as it turns out that the racist emotional response that Y experiences does not resonate with X's own emotional outlook. By contrast, others may present sentiments of disagreement that do resonate with X, causing her to revise her initial critique based on what Smith calls "a mutual sympathy" with those others.

Importantly, this is achieved through a conscious effort to engage in the sort of other-oriented perspective taking outlined in Chapters 1 and 3: one considers the psychological and cultural factors influencing the moral sentiments of others, and in so doing attempts to construct the moral perspective in question. The ease with which one can do this is dependent on whether one is able to understand and appreciate those psychological and cultural factors. In this case, despite making an effort to understand the background factors underlying Y's particular moral emotion, X is unable to construct Y's particular sentiment of approval because the psychological

and cultural factors motivating this sentiment, which X begins to understand through a nuanced consideration of Y's perspective, are racist in a way that X resists at a visceral level; X cannot empathetically construct Y's racist approval of X's own views because the underlying racism runs counter to sentiments that are deeply tied to X's moral identity. Altering such sentiments would require X to engage in profound shifts in the way that she views the world.

The takeaway from such an example is that the need to evaluate whether oneself is deserving of approval or disapproval, rather than merely approved or disapproved, is crucial in motivating the sort of wide-ranging and critical empathetic inspection of a variety of other perspectives involved in the moral question at hand, rather than basing one's judgment on a limited, biased sample of perspectives.

There is much to be said in favor of this aspect of Smith's empathy-based account of impartiality, and I hope that it is clear that it shares some central features with my own account: while acknowledging the role of sentiments in moral judgment, it encourages wide-ranging and empirically critical evaluation of one's moral views and conduct based on the empathetic examination of the moral sentiments of others. Smith's treatment of impartiality emphasizes that we cannot be complacent and unquestioning of our initial moral sentiments if we wish to be impartial, but rather must try to empathize with the sentiments of others so as to see our conduct and character through an impartial lens. In order to impartially assess whether we are deserving of the self-satisfaction that comes with feeling virtuous, we must be able to empathize with others' sentiments of approval towards our character or conduct. The above example is meant to show that it is crucial that one casts a wide net in this process of empathizing with the sentiments of others, but also to show that one's own individual moral perspective ought not be lost in this process. Empathetic moral inquiry has to walk a fine line between incorporating the sentiments

of others into one's self-evaluation and not losing one's sense of self in the process. So, one question facing an empathy-based account such as Smith's, one that grounds impartiality in empathetic consideration of the views of other people, is a question of balance: how much should one's own moral sentiments, as opposed to one's empathetic consideration of the sentiments of others, influence one's self-evaluation? In Chapter 6, I will argue that pragmatist philosophy offers a constructive answer to this question. But in the remainder of this chapter, I will discuss Smith's answer and the problems that answer presents for an empathy-based account of impartiality.

5.2: Problems for the Impartial Spectator

5.2.1: The Impartial Spectator and the Loss of Individuality

Thus far I have outlined Smith's account of the construction of the impartial spectator involved in self-assessment as an empirical, social process based on the sentiments of people that one encounters. While I have defended the empirical and social aspects of this approach, we have seen that there remains a pressing question for this empathy-based system regarding the role of one's own sentiments in relation to the sentiments of others. I will argue here that Smith's account of impartiality strips individuality from the process of self-assessment, thus blurring the distinction between one's self-assessment and others' assessment of one's moral beliefs and actions. Smith's goal in doing so is to overcome concerns regarding self-deceit and selfish biases. However, as we will see in the next section, this leaves his account open to concerns regarding intergroup bias and relativism.

In order to understand the relationship between one's individual perspective and the impartial spectator, we should begin by considering Smith's distinction between praise and praise-worthiness. Smith devotes chapter three of the third part of TMS to explaining our love of

praise and praise-worthiness and our aversion to blame and blame-worthiness. He writes that we desire not just to be praised by others, but also to be the "proper object of praise"—to be praise-worthy. Similarly, we seek not just to avoid blame, but also to avoid being the "proper object of hate" and resentment—to avoid being blame-worthy.[83]

Here again Smith is emphasizing a distinction between experiences that is based on the role of empathy in experiencing others' moral sentiments: he distinguishes between the experience of being praised and the experience of feeling deserving of praise. In order to be pleased with our conduct and character, we must feel that we in fact deserve to be praised, and we can only do so by empathetically understanding that others genuinely feel sentiments of approval towards us in the same way in which we genuinely feel sentiments of approval towards those that we admire. Our "love of praise-worthiness" (p. 114) is thus not derived from the love of praise; rather, it is derived from a desire to emulate the conduct of those we love and admire, which arises because we naturally desire to "become ourselves the objects of the like agreeable sentiments" (p. 114) that we experience in considering others that we love and admire. It is not enough for us to just be admired as these individuals are; we must feel that we share the characteristics that make us experience sentiments of admiration towards them. We must be able to empathize with others' admiration towards us, to experience something analogous to the admiration that we feel towards others that we deem praise-worthy, in order to consider ourselves praise-worthy. It is the character and conduct of others, not their fame or respect *per se*, that we recognize as the objects of our own agreeable moral sentiments. Thus, in evaluating our own character and conduct, we must attempt to empathize with those who are judging us so as to ascertain whether it is our character and conduct, and not some morally irrelevant factor

83 While there are differences between the mechanisms behind love of praise-worthiness and dread of blame-worthiness, I will deal only with praise for the remainder of this section, as I think that the points made here regarding praise can be applied to blame without losing their force.

such as beauty, charisma, or fame, that is the object of their sentiments. We recognize that we love and admire other people because of their agreeable conduct and character; we do not consider them to act agreeably or have an agreeable character because they are loved and admired. It is in this sense that Smith claims that "the love of praise seems, at least in a great measure, to be derived from that of praise-worthiness" (p. 114). This love of praiseworthiness ought to motivate us to empathize with others in the process of moral self-evaluation in order to understand whether we truly deserve their praise.

We can now consider this account of praise-worthiness in terms of the social, empirical construction of the impartial spectator described above. Smith writes that when we evaluate our own conduct to determine whether it is praise-worthy, we must do so from the perspective of the impartial spectator. It is only this impartial spectator that can properly evaluate whether our conduct or character is admirable, rather than simply admired. But the impartial spectator is empirically constructed based on the sentiments of others and thus can only tell us what is admirable in terms of the views of others. While our own sentiments certainly contribute to the construction of the impartial spectator, Smith's claim is that only the sentiments that we share with others factor into this construction; any *unique* individual perspective on what one considers admirable will be stripped away, as we evaluate "with the eyes of other people." We encounter a strange tension here in that one's love of praise-worthiness is in some sense derived from one's admiration and love for certain qualities that one finds agreeable, yet one's self-evaluation of one's own praise-worthiness does not rely on any unique individual perspective on what is admirable, but rather on the impartial perspective constructed based on the sentiments of other people. It seems, then, that the meaning of "admirable" or "praise-worthy" in Smith's system is a matter of intersubjective agreement reached via empirically amalgamating the sentiments of

others and stripping away individual biases.[84] On Smith's account, the impartial spectator is composed only of those sentiments that are shared by others, thus one's self-assessment from the perspective of the impartial spectator does not in fact rely on any aspects of one's perspective that are unique to one's self.

This conclusion, while perhaps counterintuitive, is only problematic for Smith if he is committed to a role for individuality in moral self-assessment, but in establishing the impartial spectator as the ideal judge of praise-worthiness Smith seems to be denying such a role in favor of an intersubjective perspective, with the goal of eliminating selfish biases. However, the cost of this denial is that our ideal self-evaluations will be conspicuously missing the unique aspects of our personalities, as each individual will evaluate from the same intersubjective perspective shared by others; intersubjectivity is constitutive of impartiality on Smith's account. In this way impartial self-evaluation on Smith's account is difficult to distinguish from impartial evaluation of a moral actor that is not oneself. But this is precisely Smith's point. While we lose a sense of individuality that may intuitively seem important, and that I have defended in the previous chapter when considering Bernard Williams' account of integrity, Smith takes this loss as necessary to establish a universal standard of impartial moral evaluation that is meant to avoid bias based in individual sentiments. His claim is that we avoid selfish biases only when we have

84 See Rick (2007, pp. 152-153), for a similar interpretation of the impartial spectator as an intersubjective perspective composed of shared sentiments. On Rick's interpretation of Smith,

> The content of the 'third person' perspective of the impartial spectator is constituted by intersubjective norms arrived at from within a first-person plural perspective obtained from a merging of distinct evaluative horizons. The impartial spectator's evaluative stance is impartial because it doesn't favor one side or the other but is a *shared* point of view. (pp. 152-153; emphasis in original)

The importance of shared sentiments in moral judgment is also apparent in Smith's account of general moral rules, which he takes to be established by the "concurring sentiments of mankind" (Smith, 1759/1982, p. 160), as will be discussed below. My point here is that Smith is emphasizing these "concurring" or shared sentiments as the foundation of ideal moral evaluation, including self-evaluation.

stripped away any uniquely held sentiments and have brought our sentiments into agreement with others such that we experience "mutual sympathy."

5.2.2: The Impartial Spectator, Bias, and Relativism

At this point, we are left with the claim that, for Smith, self-evaluation requires an intersubjective perspective established via an empirical, social process of empathizing with others. This perspective is stripped of an individual's unique sentiments and is constructed out of shared sentiments with others that are discovered through the process of empathetically assessing one's own conduct from others' perspectives. We can now examine whether this intersubjective perspective is capable of avoiding biases and establishing the impartiality that is Smith's goal.

While Smith argues that moral evaluations should ideally be carried out from the perspective of the impartial spectator, he recognizes that it is not always easy for human beings to adopt such a perspective. It is for this reason that he devotes significant effort to explaining self-deceit and the use of general moral rules (pp. 156-161) to overcome this deceit. Smith concedes that, "the violence and injustice of our own selfish passions are sometimes sufficient to induce the man within the breast to make a report very different from what the real circumstances of the case are capable of authorizing" (p. 157). In other words, we may sometimes convince ourselves that we have adopted the perspective of the impartial spectator when we are not in fact being impartial. Smith argues that adopting general moral rules will help us recognize and overcome such instances of self-deceit. These moral rules are adopted based on consideration of the moral behavior and judgments of others, which as we have seen is an empathetic process in which one takes on others' perspectives. As Fleischacker (2011) puts the point:

> the Smithian solution to self-deceit is not, as current catch-phrases would have it, to try to 'be myself' independently of what other people think of me. It is instead to seek my true self in the judgments, and ideally with the help, of all those other people. (p. 23)

In this section I want to examine two concerns that arise out of the question of just who exactly "all those other people" should be. In doing so I will return to some themes of Chapter 2 in order to argue that the sort of empathy-based construction of impartiality that Smith favors, in which individuality is subsumed by a perspective constructed based on the sentiments of others, is especially susceptible to problems stemming from intergroup empathy bias. The first problem facing Smith's account is this: if we construct impartiality based on empathizing with the sentiments of others, how will we avoid being biased regarding our selection of the perspectives that we empathetically consider in such a construction?

This leads to the second concern of this section. The concern is that, in stripping one's individuality from the process of impartial self-evaluation, in locating the impartial spectator in an entirely intersubjective perspective composed only of shared sentiments, Smith's view collapses into a problematic cultural relativism in which impartiality varies according to the culture in which one is situated, regardless of one's own moral convictions.

Self-Deceit and Intergroup Empathy Bias

It is significant that Smith's treatment of self-deceit is located in the section of TMS that is concerned with our sense of duty. Smith addresses the issue of self-deceit in moral evaluation, which arises when "the fury of our own passions constantly calls us back to our own place, where every thing appears magnified and misrepresented by self-love" (p. 157), by providing an account of the construction and application of moral rules according to which we can avoid the "delusions of self-love" that often cloud our evaluations of our own moral character and conduct.

Importantly, Smith argues that we feel a duty to adhere to these rules neither because of any inherent moral sense,[85] nor because our judgments of right and wrong are "formed like the decisions of a court of judicatory, by considering first the general rule, and then, secondly, whether the particular action under consideration fell properly within its comprehension" (p. 160). Rather, Smith's claim is that we construct these general moral rules based on our experience of empathizing with the moral sentiments of others in situations in which we are not ourselves the objects of approval or disapproval. According to Smith, "[i]t satisfies us that we view them in the proper light, when we see other people view them in the same light" (p. 159). As we have seen in the previous section, according to Smith, to see that others view a moral action or view in a particular light is not merely to see others praise or blame that action or view; it is to empathize with the sentiments driving that praise or blame. It is only though empathizing with these sentiments that we understand the praise or blame as relating to praiseworthiness or blameworthiness.

With this in mind, we can summarize Smith's account of the construction of moral rules as follows. We observe the conduct of others and observe how other people assess this conduct. In some cases, we observe that a particular sort of conduct is the object of sentiments of disapprobation. This is an empathetic process in which we take on the sentiments of those who disapprove of the conduct and understand that we ourselves do not wish to be the objects of such negative sentiments. As such, we resolve to avoid the sort of behavior that generates these sentiments in others; we "lay down to ourselves a general rule, that all such actions are to be avoided, as tending to render us odious, contemptible, or punishable, the objects of all those

85 See TMS (p. 158) for Smith's critique of the idea that we are "endowed with a particular power of perception" that allows us to judge the merits or demerits of moral sentiments. Smith's criticism, directed at Francis Hutcheson's moral sense theory, is that such a faculty ought to be able to more accurately assess our own sentiments, as these would be "more immediately exposed to the view of this faculty." However, the opposite seems to be the case, as we often have the most difficulty impartially judging our own sentiments.

sentiments for which we have the greatest dread and aversion" (p. 159). An analogous process leads us to construct rules according to which we ought to act in ways that generate sentiments of approval in others. In both cases, "the general rules which determine what actions are, and what are not, the objects of each of those sentiments, can be formed no other way than by observing what actions actually and in fact excite them" (p. 160). Like the impartial spectator, moral rules are constructed empirically and socially.

In terms of self-deceit, Smith's aim here is to outline a method according to which our natural tendency towards selfish biases in the evaluation of our own conduct can be overcome by constructing and applying empirically-grounded moral duties that are based on our empathetic consideration of others' judgments in situations in which we are not the objects of moral consideration. The goal is to preserve impartiality in self-assessment through a sense of duty to impartially, empathetically constructed moral rules. However, in what follows I want to make the case that the construction of these rules will not be impartial in the way that Smith hopes, in particular because of the potential for intergroup empathy bias. My aim is to show that an empirical, social, empathy-based system such as Smith's ought not appeal to empathy as a means of constructing overarching, infallible rules, as any rules that one constructs based on empathy with the views of the society in which one is situated are susceptible to partiality in the same manner in which one's own unreflective self-evaluations are susceptible to partiality. The reason for this is that the moral rules that one constructs may be based only on perspectives that are similar to one's own, and thus do not help one truly escape selfish biases in the manner that Smith suggests. In Chapter 6 I will argue that a pragmatist approach that does not utilize empathy in the service of constructing general rules is better equipped to handle these concerns of biases and can do so while maintaining a role for empathy in fallibilistic moral inquiry, but for

now it is my aim simply to draw out this pressing issue facing Smith's empathy-based account so as to highlight the need to reevaluate what impartiality ought to look like in empathy-based moral inquiry.

Recall from Chapter 2 that perhaps the most pernicious and well-documented form of bias in our empathetic abilities is bias towards favoring empathizing with those who share our own group memberships and perspectives, and against empathizing with those who do not. This is what I have been calling intergroup empathy bias. It is particularly problematic for Smith's account of the construction of moral rules for the following reason: Smith locates moral rules as originating from our experience of empathizing with the moral sentiments of those around us, but intergroup empathy bias may lead us to favor empathizing with those who already share our particular beliefs, thus leading to the construction of moral rules that only confirm, rather than challenge, certain selfish biases.

We can appreciate this point by considering historical examples in which abhorrent moral rules in certain societies were codified on the basis of general societal sentiments of approval. Consider for example slavery in the United States during the 18th and 19th centuries. Suppose a pro-slavery individual living on a plantation in Georgia in 1800 were to question his moral sentiments of approving of slavery using Smith's approach. This individual ought to construct a moral rule regarding treatment of slaves based on empathizing with how the treatment was either the subject of sentiments of approval or of disapproval in his experience. The problem is that his experience is largely limited to those who share his particular background, and intergroup empathy bias may render him more apt to empathize with those who share that background even when he encounters more diverse perspectives grounded in backgrounds that are different from his own, such as those of slaves or abolitionists.

Remember that for Smith, it is not empathy with the object of approval or disapproval that is used to construct moral rules, but rather is empathy with those who are expressing sentiments of approval or disapproval. As a result of intergroup bias, this pro-slavery individual is likely to have several experiences of empathizing with others' sentiments of disapproval of abolitionists, as he finds it easier to empathize with those who share his background and perspective. Furthermore, he does not wish to be the object of disapproval in the manner of the abolitionist, therefore, as Smith outlines, he lays down a rule such that he will avoid being such an object: he will avoid pursuing abolitionist views and conduct.

The issue here is that appealing to empathy with the sentiments of others has not enabled the pro-slavery individual to remove the self-deceit involved in justifying his pro-slavery views to himself. Rather, it has only bolstered his self-deceit, as he now will appeal to his view as being not merely a selfish bias, but rather an impartial view supported by the sentiments of others. This is a problem for Smith's account because it does not seem that the pro-slavery individual has failed to follow any of the steps outlined by Smith for empirically constructing impartial moral rules.

Now, Smith may reply that the pro-slavery rule is not an adequate rule, because it is only when a rule is "*universally* acknowledged and established, by the concurring sentiments of mankind" (p. 160, my emphasis) that it counts as a genuinely impartial rule. However, the problem is that it is unclear how one can non-arbitrarily draw the line according to which some sentiments and not others ought to be considered when appealing to "the concurring sentiments of mankind." One cannot empirically construct rules that are consistent with *all* of the sentiments of mankind, as there will always be some disagreement about a particular moral rule. It seems

that something beyond a general appeal to empathizing with the sentiments of others is needed in order to construct the sort of impartial moral rules Smith has in mind.

A proponent of empathy-based moral inquiry has two potential responses to this sort of challenge: one can either attempt to provide the additional factor needed to develop impartial rules, or one can abandon the prospect of formulating such general moral rules. I favor the latter response. Constructing rules based on empathizing with others runs the risk of codifying biases and leaving one inflexible when facing novel moral problems and perspectives, as we have seen in the above example. However, one need not abandon an empathetic method of moral self-assessment as a result. One ought to utilize empathy as a means of fallibilistic critical self-assessment, but not as a means of constructing infallible moral rules. This is the pragmatist method I will defend in Chapter 6, drawing especially on the work of Dewey and Addams. Thus, my goal in highlighting this problem for Smith's view is not to reject an empathy-based, empirical, and social approach to moral inquiry, but rather is to reject his particular appeal to empathetically constructing moral rules as the means of addressing legitimate concerns regarding self-deceit and bias that face such an approach.

Before moving on to the method that I think can address these concerns, I must first draw attention to a related issue with Smith's account in order to highlight another potential stumbling block for an empathy-based approach to moral inquiry. This is Smith's rejection of a role for individuality in the construction of the impartial spectator. While an empathy-based approach to moral inquiry that appeals to general moral rules is subject to serious concerns regarding biased formation of those rules, an empathy-based approach that rejects a role for individuality in self-assessment is subject to concerns regarding a collapse into a moral relativism in which one loses moral agency and identity.

Impartiality, Relativism, and the Loss of Moral Identity

To get at this issue, let us consider the following problem. There appear to be cases in which the public may want to say that an "impartial spectator" would claim that an individual is in fact being *too* hard on himself or herself. The issue is that these self-critical individuals' constructions of "impartial spectators" may be of much more critical judges than the impartial spectators imagined by most people. Consider, for example, an individual who donates a significant amount of time and money to charity, but who honestly believes that an impartial spectator would not deem him praiseworthy for such actions. Perhaps he believes that an impartial spectator would require that he should contribute even more time and money in order to be truly admirable or praiseworthy. The question is whether such a self-critical individual is incorrectly imagining an impartial spectator, and if so, how he will be able to determine that his conception of impartiality is too strict. Is the public correct in thinking that the impartial spectator constructed by the charitable individual is too strict, or is the charitable individual correct in thinking that the public is too lax in its conception of the impartial spectator? The underlying question is how to arbitrate between different conceptions of impartiality, and the problem for Smith's account is that it seems that we cannot do so from the perspective of an impartial spectator without falling into an infinite regress of appeals to impartial spectators: if an impartial spectator must rule on which perspective of impartiality is correct, we must then ask if *that* impartial spectator is truly impartial, and would thus need to appeal to another impartial spectator to decide, and so on, *ad infinitum.*

Perhaps the most intuitive response to this sort of problem would be to claim that it is because of the charitable individual's particular personality that he has such high standards for praiseworthiness, and that nevertheless his conduct is commendable. Thus, though he may fall

short of praiseworthiness in his own self-assessment, others can still rightly consider him praiseworthy. But given Smith's claims regarding the role of the impartial spectator in self-assessment, we should not have this kind of disconnect between self-assessment and the assessment of others if both sides are constructing the impartial spectator properly. As we have seen, self-assessment is ideally carried out by an impartial spectator that is stripped of individuality and constructed based on empathizing with the sentiments of others. For Smith, the impartial spectator is a perspective composed only of *shared* sentiments. As such, it is difficult to see how Smith can accept an explanation in which an individual can *properly* deem himself not worthy of praise while *properly* being praiseworthy in the eyes of others, as it is the intersubjective perspective of the impartial spectator that accounts for accurate judgments of propriety on Smith's view.

Smith does at times try to emphasize what Broadie (2006)[86] calls the "interiority" of the impartial spectator as a perspective that is actively constructed by an individual rather than passively accepted as a "representative of established social attitudes" (p. 181), but the discussion of Smith's view in the previous sections should leave us skeptical that Smith can in fact distance the impartial spectator from the views of society in the way that he seeks. We have seen so far that the impartial spectator is constructed empirically based on empathizing with the sentiments of others, and that "[t]he general rules which determine what actions are, and what are not, the objects of each of those [moral] sentiments, can be formed no other way than by observing what actions actually and in fact excite them." (Smith, 1759/1982, p. 160). Observing which actions excite particular moral sentiments involves empathizing with those who are judging the actions, a process that is in a significant sense not interior insofar as it necessarily involves one's experiences within a society whose members experience moral sentiments

86 Broadie focuses on the relationship between social convention and the impartial spectator on pages 181-187.

directed towards others; this includes sentiments directed towards oneself conceptualized as an other by the target of empathy. Though one could point to the role of individual imagination in amalgamating these sentiments and finding common ground, the fact remains that any common ground we find is still based on the original raw data of actual sentiments discovered empirically via our experiences of empathizing with the moral sentiments of those around us.

Thus, in the example at hand, it must be the case that either the charitable individual or those judging him have improperly constructed an impartial spectator, because the ideal impartial spectator is composed of shared sentiments, and in this case there is clear disagreement; the two opinions cannot be composed of the same shared sentiments. So, Smith may be forced to answer the above objection in the following manner. The charitable individual's standards are in fact too high. If he were to strip his own individual preferences from his self-assessment and consider his conduct from the perspective of a properly constructed inter-subjective impartial spectator, he would see that he is in fact praiseworthy. Although he considers his initial self-assessment to be impartial, it is in fact not so. The fact that the public deems him praise-worthy based on the same conduct that he is considering in his own self-assessment could be seen as evidence that he is being too hard on himself. After all, the impartial spectator should be empirically constructed based on empathizing with the sentiments of other people. The charitable individual is failing to engage in such an empathetic construction and is instead relying on his own perspective; he is failing to engage in the empirical process of seeing his conduct "through the eyes of other people" and thus is not constructing a proper impartial spectator. If he were to do so, he would effectively share the intersubjective perspective shared by those who do in fact judge him to be praiseworthy.

But this response is problematic precisely because of the notion of the empirical, socially constructed impartial spectator on which it relies. Smith's view is that while we construct the impartial spectator empirically, we must also strip the spectator of any personal biases or inconsistent perspectives. His point is that we are ideally left with a perspective that represents only the *shared* sentiments of humanity. Yet, it is unclear how this stripping process is supposed to work without running into the aforementioned worry of an infinite regress of appeals to impartial spectators as judges of which aspects to strip from our perspective, or into a problematic reliance on underlying social conventions. To see this point, let us return to the case of the charitable individual. Again, the problem in this case is that an individual's self-evaluation of praiseworthiness, carried out via his particular construction of an impartial spectator, is inconsistent with others' evaluation of his conduct. The charitable individual's conception of impartiality varies from that of the society that is evaluating him. But now imagine that there is a society in which the charitable individual's perspective is prevalent—a society that has high standards for charitable activity and expects its members to adhere to such standards. Considered relative to the views of this society, the charitable individual's self-evaluation via an impartial spectator appears to be correct. Any impartial member of *this* society would not in fact consider his conduct to be praiseworthy, as the shared sentiments of approval held by members of such a society correspond to a higher standard of charitable conduct. Yet, a society with lower standards for charity may still consider the charitable individual to be praiseworthy. So, now the question has shifted from how to arbitrate between an *individual's* construction of impartiality and a broader societal perspective, to how to arbitrate between separate *societies*' conceptions of impartiality. We began with a concern regarding individual relativism; that is, we sought to answer how an individual's notion of impartiality could be held to broader standards. Smith's

account of the construction of impartiality based on empathizing with others may initially seem to answer this concern by appealing to the ideal impartial spectator as constructed only out of shared sentiments. However, this response has now led us to concerns regarding cultural relativism in that we are left unable to arbitrate between different societies with different shared sentiments.

Smith's account lacks the tools to determine which society has constructed the correct impartial spectator. This is because in this case it is not clear on Smith's view which sentiments should be stripped away in order to construct a proper impartial perspective. We may try to find shared sentiments between the two societies, but there will always be cases in which there is definite disagreement, and it is unclear how to decide which side is more impartial without appealing to an impartial spectator to arbitrate and thus begging the question. Our criteria cannot be that the impartial perspective is one that is shared by *all humanity*, because no matter which perspective we decide on in arbitrating a disagreement, this perspective will not be shared by the society that we choose to discount; this lack of shared sentiment is why there is disagreement in the first place. It appears, then, that we are left in the same predicament at the heart of the original objection: we cannot appeal to an impartial spectator to tell us what constitutes impartiality without falling into an infinite regress or begging the question at hand based on our own underlying social conventions of impartiality.

At this point I hope to have established the following points. Smith's system overlaps in important ways with the sort of empathy-based, empirical, and social approach to moral inquiry and self-evaluation that I have defended in the previous chapters and that I will defend in Chapter 6. However, Smith's particular system faces two significant problems that need to be addressed in defending its empathy-based approach moral inquiry: (1) its reliance on the

empathetic construction of general moral rules leaves it open to concerns regarding empathy's susceptibility to bias and the possibility of codifying biases as inflexible rules, and (2) its elimination of individuality from playing a role in self-evaluation leaves it open to concerns regarding a collapse into cultural relativism. In the final chapter of this dissertation, I will address both of these concerns, drawing on insights from a pragmatist approach to inquiry. My goal is to defend an empathy-based approach to moral inquiry that shares Smith's emphasis on empirical and social reflection, but to replace Smith's approach to constructing an idealized impartial spectator via empathy with a pragmatist approach to inquiry that defines impartiality in terms of the continuous application of an empathetic, fallibilist method of moral inquiry. Unlike Smith's approach, this pragmatist-inspired approach to impartiality is based on an open-minded approach to addressing particular moral problems rather than generating general moral rules and emphasizes a significant role for both empathy and individuality in moral inquiry. I will argue that my account can address the pressing concerns facing Smith's empathy-based sentimentalism regarding bias and relativism, while maintaining Smith's general emphasis on the importance of empathy in critical moral inquiry.

Chapter 6

Impartiality as Empathetic, Fallibilistic Method: Insights from Pragmatism

It can be difficult to clarify exactly what "pragmatism" means and to list precisely what particular philosophical commitments one must hold in order to be a pragmatist.[87] There is a sense in which "pragmatist" is a historical label that groups together certain late 19th and early 20th century American philosophers such as Charles Sanders Peirce, William James, and John Dewey, who are often taken to be the paradigmatic pragmatists, but these three philosophers, as well as other philosophers of that period labeled as pragmatists, held nuanced individual views that, while overlapping with one another in a number of areas, also diverge from one another in significant ways. Furthermore, pragmatism is espoused by a number of contemporary philosophers with their own nuanced views; it is not merely a historical label.

With this in mind, it is important to make explicit just what exactly I mean by the "insights from pragmatism" referred to in the title of this chapter. My goal here is to draw on some aspects of pragmatist ethics in order to defend my account of the value of proactive empathetic engagement in moral inquiry. The pragmatist ideas I will draw on provide a way of understanding empathetic moral inquiry that emphasizes particular, concrete moral problems over rule-based moral thinking, and that emphasizes a role for individuality in the evolution of the moral self through the process of inquiry. The fundamental tenet of pragmatist ethics is a commitment to continuous open-minded experimentation in moral inquiry, and one who is committed to this tenet will be uniquely motivated to resist the empathy biases discussed throughout this dissertation, and to do so in a manner that avoids the problems facing Smith's empathy-based approach to impartiality.

87 See, for example, Lovejoy (1908).

My primary inspirations for the method of moral inquiry outlined in this section are Jane Addams and John Dewey, who I take to share pragmatist views that are especially helpful in addressing the problems discussed in the previous section. While the method that I describe is grounded in what I take to be foundational commitments of pragmatism, it is not a method that necessarily presumes pragmatism in other areas of philosophy. It is not my goal to defend the truth of pragmatism. Rather, my goal is to draw on certain foundational commitments of pragmatism to outline a method of empathy-based moral inquiry and impartiality that avoids the problems that we have seen face a Smithian sort of account. These foundational commitments are the following: fallibilism, anti-absolutism, and democracy.

In articulating how a pragmatist approach relies on proactive empathetic engagement to realize these philosophical commitments in moral inquiry, my aim is to provide a response to the problem of empathy bias in general, as outlined in Chapter 2, and to the problems facing Smith's empathy-based account in particular: a response that does not remove empathy from our moral lives, but rather seeks to utilize awareness of empathy bias to appropriately correct empathetic engagement when engaging in moral inquiry. A pragmatist approach will resist limiting empathy, because for a pragmatist empathy plays a central role in fulfilling moral inquiry's commitment to experience, democracy, and fallibilistic method. Insofar as these are good commitments to hold, then the pragmatist is right to seek to remedy empathy, rather than to limit empathy because of its susceptibility to bias; to limit empathy is to lose the benefits of empathy as a means of gathering morally relevant empirical evidence in a fallibilistic manner that benefits impartial inquiry.

To sum up, I will argue for two related claims in this section: (1) a pragmatist approach to ethics, insofar as it is committed to fallibilism, anti-absolutism, and democracy, makes empathy

a central component of moral inquiry. (2) Given the significance of empathy in pragmatist moral inquiry, the problem of empathy bias is salient for a pragmatist approach just as it is salient for Smith's account; however, the pragmatist method of inquiry is uniquely suited to remedy this problem by adjusting, rather than eliminating, empathetic engagement because it is committed to impartiality as fallibilistic method. Along with the arguments of prior chapters, insights from a pragmatist approach to impartiality in moral inquiry can effectively address the sort of general concerns held by critics such as Bloom and Prinz that were discussed in Chapters 2 and 4, as well as the concerns regarding bias and relativism facing Smith's particular sort of empathy-based account that were outlined in Chapter 5.

I will first address (1) in 6.1 by detailing each of the pragmatist commitments listed above and highlighting their connection to empathy in moral inquiry. I then address (2) in 6.2 by offering a pragmatist approach to addressing empathy bias in general and to addressing the specific problems that face Smith's view due to his commitment to general moral rules and to a lack of individuality in self-assessment. This pragmatist approach is grounded in the commitments that will be outlined in 6.1, particularly as expressed in the work of Jane Addams and John Dewey. I hope to show that the pragmatist approach can help address the general concerns of empathy bias and the specific objections I directed at Smith's empathy-based account of moral inquiry, and to do so in a manner that retains a strongly fallibilistic, experience-based, and democratic method of moral inquiry that relies on empathy.

6.1: The Role of Empathy in Pragmatist Ethics

Before examining how insights from pragmatist moral inquiry can help us address both empathy bias in general and the particular problems facing Smith's empathy-based account of impartiality, I will first examine how empathy relates to the core pragmatist commitments of

fallibilism, anti-absolutism, and democracy. My aim in doing so is to stress that pragmatist ethics, like Smith's sentimentalist ethics, and like the account of empathetic moral inquiry I have outlined thus far in the dissertation, is empirically and socially grounded and relies on empathetic engagement with others.

6.1.1: Fallibilism

Fallibilism is the view that one's beliefs are always open to revision, and that one cannot be certain that one has attained the truth. As William James (1896a) puts the point, "to hold any one [opinion]—I absolutely do not care which—as if it never could be reinterpretable or corrigible, I believe to be a tremendously mistaken attitude" (p. 14). This openness to revision is at the heart of pragmatism and is directly related to pragmatist theories of truth. While different pragmatists hold different conceptions of truth, none of which I wish to defend here, in terms of empathetic moral inquiry, the important point is the fallibilistic impulse that runs through these conceptions. For example, Peirce (1877) argues that truth is what the community of investigators will agree upon when following a fallibilistic scientific method, and James (1907) offers a fallibilistic holism[88] in which ideas "become true just insofar as they help us to get into satisfactory relation with other parts of our experience" (p. 15) and as such are always open to revision as we respond to new experiences. More recent philosophers with pragmatist leanings, including Hilary Putnam (2002)[89] and Quine, have also argued for fallibilistic approaches. The key here is that pragmatists embrace fallibilism as a methodological foundation of inquiry. Ruth

88 The fallibilism involved in James' holism shares similarities with Quine's (1953) articulation of a fallibilistic web of belief. Quine explicitly sees this fallibilistic view as "a shift towards pragmatism" (p. 340).

89 For example, Putnam emphasizes that objectivity need not preclude a fallibilistic approach. Following Dewey, he writes that

> recognizing that our judgments claim objective validity and recognizing that they are shaped by a particular culture and by a particular problematic situation are not incompatible. And this is true of scientific questions as well as ethical ones. The solution is neither to give up on the very possibility of rational discussion nor to seek an Archimedean point, an 'absolute conception' outside of all contexts and problematic situations, but—as Dewey taught his whole life long—to investigate and discuss and try things out cooperatively, democratically, and above all *fallibilistically* (p. 45, emphasis in original).

Anna Putnam (2009) nicely sums up this pragmatist emphasis on fallibilistic method when she writes that "it is not the content of the sciences that should be taken as a model for objectivity; it is their methods. Specifically, what makes for objectivity is the willingness to revise one's judgments in the face of discordant experience—that is, fallibilism" (p. 283).

Crucially, the pragmatist's fallibilistic method of revising one's judgments in the face of discordant experience applies to all judgments, including moral judgments. This point is especially clear in the work of John Dewey. Dewey (1939) makes an important distinction between preferences or desires, what Dewey calls "valuings," and value judgments, what Dewey calls "valuations." The difference is that value judgments rely on empirical evidence, they hold for particular reasons, whereas preferences or desires are mere dispositions.[90] Value judgments occur when we evaluate these dispositions through fallibilistic experimentation; we seek evidence that supports the values we hold, and depending on the nature of that evidence we may in fact change our values. Following Dewey, Elizabeth Anderson (2018) points out that this evidence may take the form of emotional response to the consequences of holding such values. Holding particular values will have particular consequences on one's behavior and engagement with the world, and the emotional responses that these consequences engender are empirical facts that can work to either prove or disprove the hypothesis that one should hold a particular value: values are empirically testable hypotheses. For example, if I claim to value wealth, but feel emotionally empty upon achieving it, or feel guilt or shame over the means of achieving it, then this is empirical evidence in favor of revising the hypothesis that I should value wealth.

The distinction between valuings and value judgments is related to Dewey's (1945) distinction between "immediate sensitiveness" and "genuine conscientiousness" in moral inquiry. He writes that:

90 See pp. 13-19.

> Perhaps the most striking difference between immediate sensitiveness, or 'intuition,' and 'conscientiousness' as reflective interest, is that the former tends to rest upon the plane of achieved goods, while the latter is on the outlook for something *better*. The truly conscientious person not only uses a standard in judging, but is concerned to revise and improve his standard. (p. 301, emphasis in original).

We can think of value judgments as involving reflection regarding the merits of valuings that arise out of immediate sensitiveness or intuitive response to some moral problem. We make value judgments when we engage in genuine conscientiousness in the consideration of a moral problem; that is, we thoroughly reflect on the consequences of holding a particular value and either endorse or revise that value as a result. This reflection is *moral* deliberation insofar as we reflect on the consequences in terms of their relation to our own self-concept; we consider how the action or view in question relates to the sort of character that we want to have. While Dewey thinks that all deliberation involves a weighing of values, it is the particular kind of value that is assessed that distinguishes moral deliberation from other deliberations:

> The value is technical, professional, economic, etc., as long as one thinks of it as something which one can aim at and attain by way of having, *possessing*; as something to be got or to be missed. Precisely the same object will have a moral value when it is thought of as making a difference in the *self*, as determining what one will be, instead of merely what one will *have*. (1945, p. 302, emphasis in original)

So, if we return to the example of a value of wealth, we can see that moral deliberation over such a value has to do not with assessing how much wealth one can actually accumulate, but rather has to do with how the accumulation of wealth reflects on one's character.

At this point we can begin to see how this pragmatist conception of value inquiry as

fallibilistic empirical inquiry points to the role of empathy in inquiry as applied specifically to moral[91] values. The empirical inquiry involved in interrogating moral values must be both imaginative and emotional, and empathy is fundamentally an imaginative and emotional process. For Dewey, "[d]eliberation is actually an imaginative rehearsal of various courses of conduct" (1945, p. 303); we imagine experiencing the hypothetical consequences of various courses of conduct and weigh the merits of these consequences against one another on the basis of our imagined experience of them. In terms of moral deliberation, the important aspect of these imagined experiences is how they reflect on our character. Making this sort of determination requires us to understand not just what particular course of affairs will result from a particular course of action or from holding a particular view, but rather requires us to understand how that course of action will affect the experiences of others, and how others would judge a particular course of action and our role in it. As Dewey puts the point:

> [I]f these consequences are conceived *merely as remote*, if their picturing does not arouse a present sense of peace, of fulfillment, or of dissatisfaction, of incompletion and irritation, the process of thinking out consequences remains purely intellectual. It is as barren of influence upon behavior as the mathematical speculations of a disembodied angel" (1945, p. 303; emphasis in original).

Yet, if this sense of peace, irritation, etc. that we experience upon reflecting on the consequences of a particular solution to a moral problem is only experienced from our own limited perspective on the problem, then we have failed to engage in genuine reflection; we remain resting on "the

91 I think that many of the points I make regarding the pragmatist imperative to empirically interrogate one's moral values via empathy apply equally to the imperative to interrogate one's aesthetic values through empathy, and that, as argued in my discussion of the imaginative approach to effortful correction of empathy bias, empathetic aesthetic engagement may indeed bolster empathetic moral inquiry, but a thorough examination of the connection between pragmatist aesthetic and moral inquiry is beyond the scope of this dissertation. For a discussion of the relationship between aesthetic and moral experience in Dewey, see Fesmire (2003).

plane of achieved goods" that comes from our particular intuitions. Moral deliberation requires us to challenge our standards, to "be on the outlook for something *better*," and we cannot engage in this critical process from within the interpretative context of the very standards we should be evaluating. Empathetically imagining how others will experience and judge the consequences of a particular moral action is thus crucial to moral inquiry in that, insofar as it is a visceral, emotional experience, it draws one out of a "purely intellectual" approach to reflection that will not influence behavior, but insofar as this emotional experience is grounded in the experiences of others, it allows one to attain the critical distance needed to assess one's own conduct and views in a manner that is capable of challenging one's preconceived moral standards.

For Dewey (1916a), "imagination is as much a normal and integral part of human activity as is muscular movement" (p. 245); imagination is an *active* experience through which we test our values by simulating the consequences, both emotional and tangible, of holding those values. Recall that I have defined empathy as the process of affectively matching with another via contextualized other-oriented perspective taking. In other words, empathy is an imaginative simulation of the emotions of another. Insofar as the emotional response to holding particular values is a key empirical fact in the testing of those values, empathy is crucial in that our capacity to empathize enables us to engage with a diverse sampling of empirical evidence. Through empathizing, we are not limited to interrogating our values based solely on our own emotional responses to holding them; rather, we are able to incorporate the emotional responses of others as relevant empirical data. And it is crucial for the pragmatist that this data is empirical; empathy allows us, at least to some degree, to *actually experience* the emotional perspective of others. In this sense empathy is a means of grounding moral inquiry in empirical method. As we have seen, the pragmatic method of inquiry is ultimately a fallibilistic scientific method.

Dewey's insight was that the scientific method can apply equally well to value inquiry while retaining the fallibilistic and empirical character that "makes for objectivity" in the sense described by Ruth Anna Putnam. For the pragmatist, impartiality (or "objectivity" in Putnam's words) is constituted by the continuous application of fallibilistic method to the problems of experience. Empathy is central to the application of this fallibilistic method when it comes to the interrogation of moral values because it is a means of supplying the empirical data, namely the moral, emotional perspectives of others, that is specifically relevant for the critical examination and potential revision of our moral values as hypotheses.

There are two forms that this empathetically gathered empirical data could take. Empathy helps us to understand (1) how our values impact the emotional lives of others and (2) how the emotional lives of others may lead them to hold values that are different than our own but are perhaps better equipped to address the problem at hand. Both of these forms of evidence are relevant to moral inquiry.

To understand the distinction, we can consider an example of a morally salient "problematical situation" (to use Dewey's language). Since 2011, due to years of civil war, approximately 5.6 million Syrians have fled Syria as refugees, and 6.6 million are internally displaced within Syria; the United Nations estimates that there are 13.1 million people in need within Syria. Half of those affected are children.[92] While some favor accommodating Syrian refugees in the United States and various European countries, others do not, citing concerns regarding national security and economic issues. There is a clear value disagreement between the two approaches to the Syrian refugee crisis, and this disagreement has concrete consequences for the refugees involved. The fallibilistic nature of the pragmatist method of moral inquiry that I have been outlining requires both sides of the disagreement to subject their values regarding this

92 These statistics are from the United Nations' Refugee Agency (2021).

issue to empirical testing. One's values must be constantly open to revision, and the urgency of the practical consequences of holding those values in this case is all the more reason to interrogate them. Again, this interrogation of values involves assessing the consequences of holding those values, and my claim is that empathy is central to this assessment of consequences from other perspectives.

The first form that empathetic interrogation of values could take in this case would involve empathizing with the refugees themselves. For example, one who is opposed to allowing the refugees to enter the U.S. can subject this value judgment to testing via reading news stories and engaging with first-hand testimonials from Syrian refugees. The aim would be to gain a nuanced understanding of the refugee perspective in order to foster some degree of accurate affective matching and thus engage in some degree of empathetic construction of the emotional toll of the refugee experience. Perhaps a better empathetic understanding of the emotional turmoil of displacement would encourage the empathizer to revise his or her values;[93] a strong response to empathetic engagement with the suffering of others can serve as empirical evidence for prioritizing helping those in need over concerns regarding terrorism and economic issues. However, it is also possible that the empathizer would maintain her initial value judgment, perhaps because of the strong pull of her other values regarding concerns of economic and national security. One always enters the process of inquiry with other values operating in the background, and it is possible that these values are strong enough to resist empathetically motivated change in this case. In such a scenario one's initial values could even be strengthened via empathetic experiment, as one recognizes that those values are able to withstand concentrated effort to empathize with those whose well-being is not served by those values. Thus, we see that

93 A number of studies by Batson (2011) seem to suggest the likelihood of this option. In the studies, participants who were encouraged to empathize with others were more likely to engage in altruistic behavior such as donating money, taking over an unpleasant task, or cooperating at a cost.

empathetic engagement allows one to realize either that one's initial values should be revised in the face of strong emotional response, or that one's initial values are strong enough to resist change even in the face of some emotional response (or lack of emotional response despite concentrated effort). In either case empathetic engagement is providing a form of empirical evidence regarding the value in question. Insofar as one is following a pragmatist fallibilistic method of inquiry, one should engage in this sort of empathetic engagement in order to subject values, as hypotheses, to empirical test, thus either strengthening or weakening the value hypothesis in question, potentially to the point of revision.

The second form that empathetic engagement may take in interrogating one's values involves empathizing with those who disagree with one's own value judgments. This is the sort of empathetic engagement that I have, for the most part, focused on defending throughout this dissertation.[94] In the Syrian refugee case, one who opposes allowing the refugees asylum in the U.S. should try to empathetically engage with those who favor doing so and *vice versa*, assuming that the views of those who do not favor asylum provide some alternative solution that falls within the horizons of compassion.[95] Empathy allows us to try on the perspectives of those whom we disagree with, to better understand the psychological and cultural factors that have led

94 An interesting problem may arise here regarding those who are especially capable empathizers, namely that one may try on a morally objectionable perspective and find it convincing enough to change one's own values simply because one has especially strong imaginative capacities. In other words, one who is an especially capable empathizer might be more easily swayed than most by certain morally deviant perspectives; the heightened capacity to experience the other's emotions will strengthen the empirical evidence for changing one's values, even if those values are seen as objectionable by less adept empathizers. For an interesting discussion of this issue, see Morton (2011). Morton argues that being a "morally sensitive person . . . limits one's capacity to empathize with those who perform atrocious acts" (p. 318). Morton's response here fits nicely with my discussion in Chapter 4 of horizons of compassion and the idea that empathy ought to be motivated by compassion and not *vice versa*. As I have stressed throughout the dissertation, the goal in correcting empathy bias is not to simply maximize empathetic engagement such that we empathize with abhorrent perspectives; rather, it is to refine empathetic capacity such that it is capable of acting in the service of compassionately motivated moral inquiry.

95 Perhaps those who oppose asylum favor some other form of intervention (economic, military, etc.) to assist the refugees. Again, we ought to ask for reasons why a given view falls within the horizons of compassion when considering whether the view is worth making an empathetic effort to understand. It is difficult to see how simply ignoring a refugee crisis, that is, offering no solution, falls within the horizons of compassion.

the other to hold different values than our own. In engaging in this process, one may find that the opponent's values resonate; through gaining a more thorough appreciation of the other's perspective one may find oneself better able to construct the other's moral emotions regarding the problem at hand, and may revise the value in question as a result. Importantly, the opponent's values will resonate because of the particular emotional character of empathetic engagement; nuanced enough empathetic engagement can allow one to feel the force of another's value commitments, and this feeling can function as empirical evidence in favor of that value. Again, this sort of engagement may not be enough to sway one's values, but in either case a fallibilist about values should seek to test values in this manner, and empathy allows one to do so. Fallibilism spurns complacency in inquiry, and, correspondingly, active empathetic engagement spurns complacency in perspective taking.

Thus, we have seen that a pragmatist-influenced commitment to fallibilistic method will lead to empathetic engagement in moral inquiry as a means of subjecting moral values to experimental test. The better we are at empathizing with involved stakeholders, the more informed our judgments about what to do in a given moral situation will be, as we will have a more robust imaginative appreciation of the motivations and consequences involved in pursuing different valuations and different solutions to the same moral problem. In empathizing with others, we fallibilistically subject ourselves to experiences that may be discordant with our preconceived values, and if these discordant experiences resonate with us as we take on perspectives based on values that differ from our own, then we can revise our own values accordingly; if they do not resonate despite our effort to engage in such an empathetic process, then the value in question has been strengthened through empirical testing, just as a scientific theory is strengthened each time it is subjected to experiment and not falsified by the data it

seeks to explain. In either case, the pragmatist's commitment to fallibilistic inquiry makes empathetic engagement a central aspect of interrogating one's moral values.

6.1.2: Anti-absolutism

The pragmatist commitment to what I will call anti-absolutism in ethics involves a rejection of universal moral rules that can be applied across the variety of unique moral problems that we encounter in our actual experience. Pragmatism rejects the idea that one can derive eternal moral principles *a priori*, then apply them to experience as problems arise. Instead, pragmatists emphasize the diversity of experience and the unique nature of the particular moral problems that we face. For Dewey, ethics involves finding solutions to specific moral "problematical situations" and each of these situations will require a solution that is particularly suited to addressing the unique nature of the problem at hand. This approach in ethics stems from the pragmatist's emphasis on use rather than correspondence. Again, I do not wish to defend general pragmatist theories of truth, but this emphasis on use rather than correspondence in the moral realm is helpful in considering how we ought to approach moral inquiry in an impartial manner. For the pragmatist, moral truths are true because they *work*, not because they correspond to any objective reality that exists outside of human ends. William James (1891a) summarizes this anti-absolutist approach thusly:

> There is no such thing possible as an ethical philosophy dogmatically made up in advance. We all help to determine the content of ethical philosophy so far as we contribute to the race's moral life. In other words, there can be no final truth in ethics any more than in physics, until the last man has had his experience and said his say. (p. 184).

James' point is that the test of truth in ethics is experience itself, not correspondence to *a priori* universal moral truths. This pragmatist idea again places empathy in a crucial role for

moral inquiry. If we must look to experience to understand ethics, then we ought to look to empathetic experience in particular. Pragmatic anti-absolutism in ethics encourages us to seek out a wide variety of potentially morally salient problems and perspectives, rather than look to *a priori* principles to reveal one set of relevant moral problems and one set of correct responses to those problems. Empathy provides us with a tool to step outside of our own concerns and access this wide variety of moral perspectives, and to understand those perspectives as legitimate. If, in empathizing with another, we experience his or her perspective as morally agreeable, then that is enough for that perspective to count as morally agreeable; we will of course need to evaluate the other's valuations in terms of how they work out in experience and how they accord with our own values, but this Deweyan process of evaluation is an experiential and imaginative process of relating another's values to our own, not a process of evaluating whether another's view corresponds to some objective moral facts about the world. In moral inquiry, experience itself, insofar as it follows the fallibilistic pragmatic method of open-minded engagement and criticism, *is* justification, and this sort of experience is what empathetic engagement facilitates. Empathetic engagement actively generates a wide variety of imaginative experiences, each of which is evaluated on its own experiential terms in order to understand the diversity of stakeholders involved in the moral problem at hand. It is in this sense that Dewey (1945) refer to empathy[96] as the "animating mold of moral judgment," without which there would be no "material with which to deliberate" (pp. 269-270). Our direct empathetic experience of others' values and emotions is what drives us to critically reassess our own values. Anti-absolutism shifts the focus of ethics away from abstract *a priori* reasoning and towards an understanding of others' moral psychology, while empathy provides us with a means of accessing (i.e., experiencing the valuings of others) and assessing (i.e., engaging in Deweyan "valuations") the moral psychology

96 Dewey, like Hume and Smith, uses the term 'sympathy' to refer to perspective-taking.

of others. As Dewey (1945) writes, empathy "furnishes the most efficacious *intellectual* standpoint… Through sympathy the cold calculations of utilitarianism and the formal law of Kant are transported into vital and moving realities" (p. 270, emphasis in original). For the empirically minded pragmatist, it is this "vital and moving" *experience* that is the engine of moral inquiry.

6.1.3: Democracy

The pragmatist commitment to democracy is, as Ruth Anna Putnam notes, a commitment to democracy in a "wide sense" that includes not just the political system of democracy, but "also social, liberal, and pluralistic democracy" (p. 278). Like the pragmatist commitment to fallibilism, the pragmatist commitment to democracy is a function of the underlying scientific methodological framework that drives pragmatist inquiry in every area. As Putnam writes, "by analogy with the sciences and the arts, we may say that societies will flourish and permit their members to flourish if they permit the free exchange of ideas, including particular ideas about the organization of society itself" (p. 287). Just as we cannot ignore the evidence of experiment in testing a scientific theory, we cannot ignore the evidence of democratic experience in testing our values. To dogmatically ignore the ideas of others, to limit the free exchange of ideas, is to violate the pragmatic rule that Peirce (1898) declares "deserves to be inscribed upon every wall of the city of philosophy: *Do not block the way of inquiry"* (p. 48; emphasis in original). At its core, pragmatist inquiry is about experimentation, and when it comes to the interrogation of values, democracy is the surest way to facilitate experimentation, as the democratic society is one in which its members are exposed and open to new and unfamiliar ideas, and in which they are free to express and test their own values against the experiences of others.

Empathy will play a central role in such a democratic society. In order for a true *exchange* of ideas to occur regarding values, one has to step outside of one's own perspective and engage empathetically with the perspective of those on the other end of the exchange. Recall that a key component of empathy is other-oriented perspective taking, which is distinct from self-oriented perspective taking. It is not enough to imagine the other's perspective from one's own point of view; one ought to consider the other's perspective from within the other's point of view in order to appreciate the value system within which the other is operating. However, it is crucial for pragmatists that individuality is not lost within the democratic society, as individuality accounts for the diversity of ideas that is needed to furnish the imaginative process of deliberation. As noted in the discussion of compassion from the previous chapter, empathy is a means of appreciating the individuality of others, of truly validating the significance of their individual perspectives through making the effort to inhabit those perspectives, rather than merely judging their perspectives from the vantage of one's own framework. Dewey (1888/1993) writes that democracy "is the form of society in which every man has a chance and knows that he has it—and we may add a chance to which no possible limits can be put, a chance which is truly infinite, the chance to become a person" (p. 63). A willingness to empathetically engage with others drives the availability of this "chance" for all those involved in the democratic society. To empathize with another is to seek to remove some of the limits of one's own perspective in subjecting the other's values to test; it is to recognize the other as what Dewey calls "a personality with infinite capacities" (p. 65) worthy of consideration.

6.2: Addressing Bias and Maintaining Individuality

6.2.1: Empathy Bias and Addams' Social Ethics

Thus far I have argued that the core pragmatist commitments to fallibilism, anti-absolutism, and democracy ought to motivate us to utilize empathetic engagement in the process of impartial moral inquiry. The general claim is that empathy is a means of subjecting one's values to experimental test, thus insofar as one seeks to adhere to this sort of impartial, pragmatist-inspired experimental method in ethics one will seek to empathize in the process of moral inquiry. But here is where the problem of empathy bias arises: as we have seen, research on empathy suggests that we are subject to subconscious biases that limit our capacity to empathetically engage with those who are not like us, those whom we consider to be members of outgroups. The problem is a familiar one by this point. If empathy is a key component of moral inquiry, but we are subconsciously biased towards empathizing with others who are like us, then it seems that we may have an unfortunate problem of confirmation bias in our moral inquiries: we will be more likely to consider evidence that confirms our own values because we will be more likely to engage empathetically with members of our ingroups that share those values. Our empathy bias renders us psychologically less capable of considering *all* the relevant evidence (i.e., the perspectives of others who are not like us) when engaging in moral inquiry, thus the pragmatist goal of impartiality through method seems to be in jeopardy, as the method may itself be partial.

In articulating a pragmatist response to this sort of challenge, it will be helpful to consider the work of Jane Addams, specifically Addams' (1902/2005) conception of "social ethics." Addams' social ethics embodies the pragmatist commitments discussed in 6.1 and thus makes empathy a central component of moral inquiry. Furthermore, it is a useful example of how insights from pragmatist ethics can help us address the problem of empathy bias without sacrificing the benefits of empathetic engagement in moral inquiry. The key, as with all of

pragmatist philosophy, lies in Addams' focus on a method grounded in experience, specifically the fallibilistic, anti-absolutist, and democratic method that I have outlined in the previous section. A pragmatist is committed to this method above all else; the method itself is in fact the means of obtaining impartiality for the pragmatist. Because, as I have argued in 6.1, this method utilizes empathy when applied to moral inquiry, insofar as the pragmatist will retain the method, she will seek to adjust her empathetic capacity to compensate for bias, rather than turn away from empathy as Bloom and Prinz would advocate, because to turn away from empathy is to block the method of inquiry that is the foundation of the pragmatist approach.

For Addams, "a standard of social ethics is not attained by traveling a sequestered byway, but by mixing on the thronged and common road where all must turn out for one another, and at least see the size of one another's burdens" (p. 2). In true anti-absolutist pragmatist fashion, Addams argues that ethics cannot be conducted outside of particular experiences of what she calls "perplexities."[97] These perplexities are instances in which one's preexisting values are inadequate to understand or address a particular problem. Thus, in order to adequately solve the problem, one must undergo the empathetic process of stepping outside of one's own perspective and understanding the problem from the perspective of others who are involved. As such, in order to engage in moral inquiry, one must make a conscious effort to place oneself in the midst of perplexities and employ empathetic engagement with members of out-groups to seek solutions.[98] There is all the more impetus to place oneself in these situations once one becomes

97 Addams' use of "perplexity" can be seen as analogous, though not equivalent, to Dewey's conception of "problematical situations" that trigger inquiry. Dewey (1910, 1916b) also discusses "perplexities." For example, he writes that, "demand for the solution of a perplexity is the steadying and guiding factor in the entire process of reflection" (1910, p. 342). According to Dewey (1916b), we solve a perplexity by "conceiving the connection between ourselves and the world in which we live" (p. 354). Addams' approach to ethics can be seen as employing empathy in the understanding of the values of others as a crucial part of "the world in which we live", a part that we need to work to connect with our own experience.

98 *Democracy and Social Ethics* largely focuses on specific instances of this idea in application, including charity workers entering an unfamiliar community (pp. 11-35), and political organizers seeking to change the political

aware of the pernicious effect of empathy bias on the scope of moral inquiry. For Addams, the starting point of social ethics is recognition of the limits of one's own perspective and a desire to widen those limits as much as possible through engagement with others. In placing oneself in a diverse variety of perplexities, the goal is not to apply one's own set of values in a uniform fashion so as to solve all problems according to that particular set of values. Rather, the goal is to widen the scope of one's moral imagination by consciously seeking out and directly engaging with those whose values and life experiences differ from one's own.

Addams recognizes that "much of the insensibility and hardness of the world is due to the lack of imagination which prevents a realization of the experiences of other people" (p. 3), but her solution is not to shift focus away from imagination, but rather to consciously place oneself in situations in which one's moral imagination has the opportunity to be empathetically stimulated via direct, immersive engagement with individuals from a variety of backgrounds. For Addams, ethics involves an *obligation* to seek out the outgroup and to address perplexities by consciously attempting to take on their perspectives. This is foundational to the pragmatist method by which we achieve impartiality in moral inquiry. Thus, it is the conviction that moral inquiry must be impartial that ought to lead us to empathize; it is not that empathetic engagement leads us to be impartial. Crucially, this conviction falls out of Addams' commitment to the empirical, democratic, and fallibilistic pragmatist method. She writes that, "there is a conviction that we are under a *moral obligation* in choosing our experiences, since the result of those experiences must ultimately determine our understanding of life" (p. 3, emphasis mine). Rather than reject empathy's relevance to moral judgment, Addams recognizes that accepting the pragmatist method as one's guide to moral inquiry leads to a moral responsibility to *choose*

---

environment of a community of which they are not a part (pp. 98-102). The underlying theme is that these sorts of perplexities are only solvable when the outsider works with those involved from within an empathetic understanding of their perspectives.

diverse experiences and empathize with diverse perspectives. Dewey echoes this sentiment when he writes:

> We shall have to discover the personal factors that now influence us unconsciously and begin to accept a new and moral responsibility for them… so long as we ignore this factor, its deeds will be largely evil, not because *it* is evil, but because, flourishing in the dark, it is without responsibility and without check" (1916b, p. 327).

Pragmatists should embrace evidence of empathy bias as helpful to their ethical program. This evidence should alert us to harmful complacency, spur us to prevent empathy bias from "flourishing in the dark, and force us to take moral responsibility for the ways in which we choose our experiences so as to remedy this fundamental problem. It should not, however, lead the pragmatist to side with critics like Bloom or Prinz's approach to the problem. The pragmatist will not appeal merely to the utilitarian cost-benefit analysis Bloom offers as alternatives to empathy; pragmatists will take an unchecked utilitarian approach to be too inflexible to deal with the diversity of perspectives involved. Furthermore, the pragmatist may point out that Prinz, in appealing to the power of emotions such as guilt and anger, fails to appreciate the social, democratic construction of these emotions through empathetic engagement with others.

In sum, the pragmatist response to the general problem of empathy bias is that it would be a mistake to tamp down empathy; doing so would only block the way of moral inquiry. Rather, we should seek to compensate for empathy bias by following Addams and Dewey in recognizing that the pragmatist method yields a moral obligation to actively seek diverse experiences, including diverse experiences of empathetic engagement with others. Diverse and nuanced empathetic engagement qualifies as what Addams calls "genuine experience": a gathering of the relevant experiential data to address the moral perplexities that we encounter in

our actual lives; as Addams writes, “we do not believe that genuine experience can lead us astray any more than scientific data can” (p. 2). This approach exemplifies pragmatist philosophy’s capacity to embrace and productively engage with contemporary empirical work in moral psychology, to recognize the problems facing moral inquiry as a result of bias, and to look to Addams’ method of immersing oneself in foreign moral perplexities as a means of correcting bias through social experience.

We can now turn to the problems that plagued Smith’s empathy-based account in order to show that the pragmatist insights outlined thus far can address these concerns while retaining the value of empathy in moral inquiry.

6.2.2: Principles, Rules, and Empathetic Experimentation

The first problem facing Smith’s account was that its appeal to the empathetic, social development of moral rules failed to address concerns regarding intergroup empathy bias. If moral rules are constructed according to empathizing with the moral perspectives of others, we run the risk of codifying biased moral views as moral rules because of our biases that favor empathetic engagement with certain social groups over others. In what follows I want to highlight Dewey’s (1945) distinction between principles and rules as a means of avoiding this problem. As we shall see, Dewey’s view does not sacrifice the social, empirical approach that both pragmatists and Smith favor, but in locating impartiality in a fallibilistic and continuously critical method rather than in an effort to fix moral rules, it avoids the problems regarding in-group bias that plague Smith’s account.

Dewey’s distinction between principles and rules is meant to preserve the value of general moral views while avoiding fixing those views in a manner such that they cannot be revised and refined based on the experience of particular moral problems. Whereas principles are

flexible tools that can help us identify and reflect on the salient aspects of a moral problem, rules are ready-made prescriptions for action that only allow for one correct solution and thus shut off potentially productive avenues of inquiry. Dewey puts the point as follows:

> Now a genuine principle differs from a rule in two ways: (a) A principle evolves in connection with the course of experience, being a generalized statement of what sort of consequences and values tend to be realized in certain kinds of situations; a rule is taken as something ready-made and fixed. (b) A principle is primarily intellectual, a method and scheme for judging, and is practical secondarily because of what it discloses; a rule is primarily practical. (1945, p. 305)

As we have seen, Smith's account of moral rules is mostly in line with what Dewey calls principles according to (a). That is, Smith is explicit that moral rules are not ready-made but rather are derived from experience, particularly the experience of empathizing with the moral sentiments of others. Smith and Dewey agree about the social and empirical nature of what Dewey calls principles and what Smith calls rules. However, it is (b) above that separates Dewey's account in a helpful way from Smith's.

Smith's goal in providing an account of the formation of moral rules is to highlight a means out of the self-deceit that can motivate and justify partial, immoral actions because of the emotional strength of self-love. For Smith, we must form moral rules so as to have ready-made prescriptions to follow in the face of the strong emotional experience of self-love and self-deceit. For example, in discussing a man considering violent revenge against an enemy in response to "no more than a slight provocation," Smith writes that, "reverence for the rule which past experience has impressed upon him, checks the impetuosity of his passion, and helps him to correct the too partial views which self-love might otherwise suggest, of what was proper to be

done in his situation" (p. 161). The point here is that a moral rule regarding a just response to being wronged motivates the man in question to *act* appropriately despite his initial "impetuous" passions.

Now, Dewey's account is not opposed to principles enabling action—a principle is practical, but it is practical *secondarily*. As such, Dewey's pragmatist account of principles could endorse the vengeful man's application of a principle derived from experience regarding the appropriate response to being wronged in such a manner. The man carries a principle of justice according to which wrongs should not be avenged in a disproportionate manner. He considers the particulars of the case at hand, considers the consequences of responding in various ways and how these consequences will be viewed by others. Again, note that empathy is central in this process insofar as one needs to understand how others will view one's moral actions, and how others have viewed similar actions in the past. In this way the principle of justice encourages the man to direct reflection towards certain kinds of consequences; the principle serves as a tool to direct reflection towards the relation of the wrong to his potential responses and how such a relation will be viewed both by others and by himself after the heat of the moment. Upon considering these consequences for this particular situation, employing empathy in the process so as to imagine how others will judge his response, the man chooses to act in a certain manner. Thus, the principle is practical secondarily in that it directs the reflection and deliberation on which an action is based, but does not directly prescribe a particular action.

Dewey's account is not opposed to the use of general moral views to correct for self-deceit, but it is opposed to such general moral views directly prescribing action rather than catalyzing reflection on relevant potential consequences of the particular situation at hand. On the pragmatist view, it is not the action of the man that is problematic in this case, nor is it the

appeal to general moral ideas derived from experience; rather, it is the "reverence" for such rules that Smith emphasizes that is problematic. In the example at hand, a pragmatist method that employs principles as tools to direct moral inquiry to relevant considerations in this particular situation may very well end up leading the man to the same solution as Smith's method of acting directly out of a reverence for moral rules derived from past experience, but this will not always be the case, and it is crucial to emphasize the subtle distinction between these two approaches. The distinction is that, while Dewey's pragmatist method employs principles as tools to direct inquiry in a *forward-looking* manner, to direct action based on reflection on the particular aspects and potential novel consequences of the situation at hand, Smith's account is *backward-looking* in that it emphasizes strict adherence to rules that have been codified based on prior empathetic experience with others.

In order to flesh out this distinction and its relation to the problem of empathy bias, we can again consider the example of an anti-abolitionist who codifies his support of slavery as an appropriate moral rule based on empathy with other anti-abolitionists. We can now ask how Dewey's pragmatist method of employing principles as tools differs from Smith's method of acting based on reverence for empathetically pre-established moral rules in this case. We want to be able to label the anti-abolitionist's support of slavery as decidedly not impartial, but as we saw in Chapter 5, Smith's method seems to be at a loss to do so. This is because the anti-abolitionist has followed Smith's method of constructing moral rules based on empathizing with the sentiments of others. The problem is that empathy bias has led the anti-abolitionist to empathize with those who already share his anti-abolitionist view. Regarding our moral judgments, Smith tells us that, "[i]t satisfies us that we view them in the proper light, when we see other people view them in the same light" (p. 159). The problem is that Smith's account does

not provide us with the tools to determine *which* people we ought to consider and thus is at a loss to condemn the sort of rule that the anti-abolitionist has established via empathetic biases. While our approval of certain actions is bolstered when "we hear every body around us express the same favourable opinion concerning them" (p. 159), this is actually problematic when everybody around us shares our biases. This problem can become exacerbated given Smith's claim that we lay down rules such that "every opportunity of acting in this manner is carefully to be sought after" (p. 159). Thus, we may end up with codified biased rules for action to which we attribute impartiality and which we take "every opportunity" to pursue out of "reverence."

Although Smith shares the pragmatist's empiricist bent in his emphasis on the formation of general moral rules from a foundation of particular experience, his account departs from the pragmatist approach in his lack of emphasis on cases in which those moral rules ought to be challenged. That is, in emphasizing only the role of *prior* experience in constructing moral rules, Smith does not address the possibility that such prior experiences may be ill-equipped to handle novel moral problems and that such rules should be open to revision. For Smith, impartiality is achieved through combating selfish biases via moral rules established based on shared sentiments with others. Once we codify such rules, we act in accordance with the action that they prescribe. Yet, as we see in the anti-abolitionist case, these shared sentiments may prescribe an action that we ought not simply adhere to, that we ought not have reverence for, and that we ought to criticize and revise.

By contrast, for the pragmatist, reverence is not for prescribed actions based on prior experience but rather is for adherence to a method in which we remain open to revising our views based on novel consequences and previously unconsidered morally salient aspects of a situation that may be similar, but importantly different from our past experience. So, while a

pragmatist method will not alter our psychology such that the pull of empathy bias is never a factor, adherence to the pragmatist's fallibilistic and anti-absolutist method will deter us from allowing that bias to codify ready-made and inflexible moral rules. Pragmatist moral inquiry is based on openness to adjustment in the face of new problems and on a commitment to engaging with diverse perspectives so as to appreciate what those problems are. As such, according to the pragmatist method, the anti-abolitionist ought to challenge his empathetically constructed moral views regarding slavery, and he ought to do so precisely by making an effort to empathetically engage with those who differ from him, i.e., with slaves and with abolitionists. For the pragmatist, the solution to the problem of empathy bias is not to abandon Smith's view that impartiality must in some sense be socially and empathetically grounded; to think otherwise would be to abandon the anti-absolutism that is foundational to pragmatist inquiry. Rather, the solution is to amend Smith's view such that impartiality is constituted by continuous application of the fallibilist method of inquiry itself, and not by universalizing any solution that this method might lead one to favor in one particular situation. Thus, the key difference between Smith's account of moral rules and Dewey's pragmatist account of moral principles is Dewey's rejection of the spirit of reverence for rules that runs through Smith's account.

It is reverence for critical pragmatist method that can truly combat empathy bias. Dewey's pragmatist method locates impartial morality directly in one's openness to growth and change. We cannot pursue such growth and change and remain complacent regarding any general moral views we acquire. As such, we cannot be moral without consciously pursuing the behaviors that remedy empathy bias. As Dewey writes:

> Indeed, we may say that the good person is precisely the one who is most conscious of the alternative, and is the most concerned to find openings for the newly forming or

> growing self; since no matter how "good" he has been, he becomes "bad" (even though acting upon a relatively high plane of attainment) as soon as he fails to respond to the demand for growth. Any other basis for judging the moral status of the self is conventional. In reality, direction of movement, not the plane of attainment and rest, determines moral quality (1945, pp. 341-342)

6.2.3: Individuality, Impartiality, and the Evolution of the Moral Self.

The second problem facing Smith's system arose because the individuality of the moral self is stripped away in favor of shared sentiments in the process of empathetically constructing an impartial spectator. The concern was that this leaves the individual in a culturally relativistic moral framework in which the impartiality of her views is defined *only* in terms of their accordance with the views of the society in which she is situated, and she is left with no means of appealing to her own unique convictions to challenge that particular society's conception of impartiality.

I want to highlight a pragmatist conception of the relation between the moral self and society in order to address this concern. On the pragmatist view the moral self is defined in terms of its relation to society. However, it is not defined by mere shared sentiments with others, but rather by its *evolution* within a social context, and this evolution necessarily involves a role for individuality in challenging established societal norms. Whereas Smith's system subjugates unique individual perspectives to socially shared sentiments, the pragmatist method emphasizes the role of individuality in the critical, imaginative reflection involved in spurring moral development. The individual is still inextricably social on the pragmatist account, but it is not complacent. Furthermore, moral growth at a societal level arises out of the conflict between individuality and convention.

A pragmatist account of the moral self is able to avoid the problem of cultural relativism because, unlike Smith, the pragmatist does not locate impartiality in a set of shared sentiments; rather, for the pragmatist, impartiality arises out of the application of a fallibilistic method of inquiry. The problem for Smith's account arose because we were left without a means to arbitrate between different societies' varying conceptions of impartiality. We could not appeal to Smith's impartial spectator to solve the problem because the impartial spectator is constructed out of shared sentiments, but the very question at hand, namely which shared sentiments dictate impartiality, cannot be answered from the perspective of a particular collection of shared sentiments without begging the question. In addition, we cannot appeal to any uniquely held moral convictions to settle the dispute on Smith's account because such an appeal, given that it is based on sentiments not shared with the rest of society, fails to meet Smith's criterion of impartiality.

The pragmatist solves this problem by shifting the question of impartiality away from focusing on which set of shared sentiments is the impartial one and towards a focus on whether or not one's own moral beliefs can continue to withstand a tribunal of intelligent social critique. When addressing a particular moral problem, one ought to advocate for the solution that one finds impartial, however, as Dewey notes, "[i]n asserting the rightfulness of his own judgment of what is obligatory, he is implicitly putting forth a social claim, something therefore to be tested and confirmed by further trial by others" (1945, p. 252). Impartiality for the pragmatist lies not in generating the appropriate sentiments based on shared sentiments with others, but rather lies in subjecting one's own sentiments to social critique in a manner such that one is willing to alter one's belief in the face of relevant evidence but not merely because those sentiments are not shared by all within a society. For a pragmatist, impartiality is a mindset in which one remains

open to moral growth, but this often means challenging societal norms rather than conforming to them. Whereas Smith sought impartiality in the construction of an impartial spectator built on shared sentiment and stripped of individual concerns, the pragmatist defines impartiality in terms of subjecting individual concerns to continuous test, encouraging individuals to propose non-conformist ideas but nevertheless locating impartiality in a continual willingness to revise or abandon such ideas in the face of new or unforeseen challenges. Smith's account aims to *settle* impartiality in the sentiments of others, whereas for the pragmatist impartiality is not found in a fixed perspective; it is found in terms of an openness to the continuous evolution of the moral self that is often catalyzed by independent resistance to societally settled moral ideas.

The problem with Smith's account is its emphasis on an *ideal* impartial spectator. Striving for an ideal spectator implies that we can reach some fixed and infallible perspective on what the impartial solution to a given problem is. Yet this process of striving for an ideal is necessarily empirical for Smith. He describes a process by which human beings construct the impartial spectator, but then act as if this impartial perspective is an immutable ideal that they have discovered, rather than constructed themselves. It is the tension between an abstract ideal of a perfectly impartial perspective and the empirical fact that we must try to attain this perspective by empathizing within a limited social context that leads to the problem of a loss of individuality and a collapse into cultural relativism. Because one aims at an ideal impartial perspective which can be shared by "all mankind" one is motivated to suppress individual concerns in favor of shared sentiments. However, because one's only means of constructing an impartial spectator are empirical and social, and because of the empirical fact that certain views simply will not be shared by "all mankind", the best that one can do is to construct a spectator based on the shared sentiments of some particular subset of society.

By contrast, the pragmatist approach is to locate impartiality in a recognition that one cannot reach such an infallible ideal and thus must always be open to critical revision. Because there is no goal of an ideal impartial spectator based on the shared sentiments of others, one need not necessarily subjugate individuality in the moral realm. One need not necessarily seek to mold one's sentiments to be in accordance with the sentiments of others, but rather may seek moral progress by appealing to convictions not shared by the society at large. This is not to sav that the individual is the final judge of right and wrong; rather it is to say that the individual plays a proactive role in the social development of morality. Independent thought is an essential driver of fallibilistic moral inquiry in the same way that it is a fundamental driver of scientific inquiry. As Dewey puts the point:

> Independence of character and judgment is to be prized. But it is an independence which does not signify separateness; it is something displayed in relation to others. There is no one, for example, of whom independent inquiry, reflection, and insight are more characteristic than the genuine scientific and philosophic thinker. But his independence is a futile eccentricity unless he thinks upon problems which have originated in a long tradition, and unless he intends to share his conclusions with others, so as to win their assent or elicit their corrections. (1945, p. 248)

In terms of moral inquiry, the point is that independent insights are still situated within the context of socially defined moral problems and are still accountable to social assessment, but that impartiality need not require that one act or judge based only in shared sentiments.

Thus, the pragmatist does not disagree with Smith that moral inquiry is an empirical, social process. Furthermore, as argued in 6.1, the pragmatist, like Smith, embraces empathy as a central component of this process. The difference is the role that empathy plays as an evidence-

gathering tool for Smith and for the pragmatist. Whereas Smith's claim is that empathy is used to gather evidence of others' sentiments such that an ideal impartial perspective is constructed out of shared sentiments, the pragmatist ought to employ empathy to maintain a fallibilistic openness to new evidence, to remain open to new avenues of moral inquiry. So, while empathy is used to gather evidence on both accounts, that evidence is put in the service of fixing an ideal impartial perspective based on shared sentiments for Smith, while for the pragmatist empathy is used to "elicit corrections" to one's independent approaches; that is, for the pragmatist, empathy is a feature of the fallibilistic method of challenging independently purposed solutions with the goal of forward-looking moral development, not a means of constructing and cementing an ideal perspective capable of infallible judgment.

Peirce (1877) writes that "[b]elief does not make us act at once, but puts us into such a condition that we shall behave in some certain way, when the occasion arises. Doubt has not the least such active effect, but stimulates us to inquiry until it is destroyed" (sec. III). In the moral realm, we can think of empathy as a means of generating the doubt that stimulates us to inquiry regarding our own moral beliefs. As Peirce notes, "the mere putting of a proposition into the interrogative form does not stimulate the mind to any struggle after belief. There must be a real and living doubt, and without this all discussion is idle" (sec. IV). Empathy enables us to experience a real and living doubt regarding our own views in a manner that we cannot achieve by interrogating our views only from within our own perspective. When considering our conduct or views from the perspective of another and empathizing with the other's sentiments, we may be led to question our own sentiments if they are not in accordance with those of the target of empathy. However, we ought not *necessarily* fix our moral beliefs by striving to alter our sentiments so as to be in harmony with those of others, to seek out Smith's "pleasures of mutual

sympathy." The burden is not always one-directional in the sense that I must alter my sentiments such that they are shared by others. Sometimes empathizing with others allows us to see aspects of the *other's* moral sentiments that we think should be adjusted; we identify a disharmony of sentiments between ourselves and the other, but we identify the other's sentiments rather than our own as the target of adjustment. Thus, empathy gathers evidence, though that evidence is not always put in the service of self-adjustment but rather is sometimes put in the service of critiquing others. Our individuality remains a key feature of moral inquiry, but inquiry is impartial only insofar as our individual beliefs are open to revision based on empathetically gathered evidence regarding the sentiments of others. Remaining open to revision is a far different mindset than striving to fix moral beliefs through shared sentiments. Allowing shared sentiments to dictate moral beliefs is to fall into what Peirce calls "the method of authority" in fixing beliefs in the moral realm: beliefs are settled by the authority of established societal norms rather than by intelligent criticism of those norms. Remaining open to revision is consistent both with sharing some sentiments generally held by members of one's society and with remaining open to recognizing that some sentiments shared by the members of the society in which one finds oneself ought not to be shared. The key is that such recognition is impartial only when effort has been made to empathetically gather evidence regarding the other's perspective; this evidence can tell us about the other's need to change just as it can tell us about our own need to change. We must be open to both possibilities.

Yet practical considerations are such that in most cases we must act or hold moral beliefs without empathetically gathering *all* the evidence on relevant perspectives on the issue. Moral dilemmas are often exactly the sort of problems that James (1891b) had in mind when writing of the need to commit to action "in advance of the evidence." They are "forced" in that one must

make a decision (and choosing to refrain from action counts as a decision), they are "momentous" in that moral decisions matter deeply to us, and they are "live" in that the answer is not readily available through purely intellectual problem-solving. There are cases in which we cannot simply remain agnostic in the rightness or wrongness of an action until sufficient evidence is presented, because we are faced with the practical need to act in what we consider to be the most morally appropriate manner in a specific problematical situation. Consider, for example, Sartre's[99] well-known case of Pierre, who faces a dilemma regarding whether he should join the Resistance effort in France during World War II or remain home to take care of his aging mother. Pierre must act, but he does not and cannot know that one particular course of action is the right one prior to acting. No amount of empathetically shared sentiments will be able to answer this question in an impartial manner for Pierre, as this situation is uniquely related to his identity: it is a question about who he is and wants to be as human being. The impartial spectator cannot tell Pierre how to act, but he must act.

The pragmatist point about this case is twofold. First, the individual is the ultimate locus of action. Empathy remains helpful in gathering evidence regarding the decision; Pierre ought to try to empathize with Resistance members, with his mother, with citizens affected by the war, etc. But no amount of empathetic consideration will tell him what the correct course of action is prior to action. Empathy gathers evidence, but it is the individual who must weigh and act on that evidence.

Second, although Pierre must act "in advance of the evidence" in the sense that he will not know whether he will come to regret the decision for whatever reason, he remains impartial only insofar as he remains open to the possibility of such regret. That is, in pursuing a particular course of action, he does not rule out the possibility that he could come to find out that this was

99 My treatment of Sartre's example in terms of James' argument draws on Putnam's (1992) helpful discussion.

the wrong decision. Crucially, maintaining this openness involves maintaining an openness to empathize with those affected by the decision. Pierre must act before all the evidence is in, but he remains impartial insofar as he remains open to continuing to try to gather that relevant evidence via empathizing with those who disagree with his choice and with those who were affected by his choice, with the mindset that his belief may or may not change as a result.

Empathy allows the sentiments of others to become factors in one's individual choice to act or judge a certain way, but ultimately this remains an individual choice, and we must accept that we will be unable to accommodate all the relevant perspectives in many situations before acting. Nevertheless, impartial evaluation of our own conduct requires maintaining a fallibilistic openness to continuing to employ empathy in the gathering of relevant evidence with the mindset that this aids the evolution of the moral self that may face similar problems in the future.

In sum, the problem with Smith's account is not its focus on empathy or on the social and empirical aspects of moral development; the problem is that Smith focuses on a goal of fixing an ideally impartial perspective, and focusing on such a goal can lead to complacency once we believe we have attained it. Once we think we have arrived at the ideal perspective of an impartial spectator, we are apt to close empathetic avenues of inquiry with those who disagree. On the other hand, if we recognize that such an ideal is unattainable, we remain open to empathizing with others. Understanding that an ideal impartial spectator is impossible forces us to continue the process of inquiry rather than allow biases to seep back in as we assume that we have settled on an ideal impartial perspective. Impartiality lies in the continuous process of inquiry, not in the discovery of a static ideal.

The cultural relativism objection is problematic for Smith's view because of his account's insistence that we ought to make moral judgments from an ideally impartial set of shared

sentiments, when in fact no such set exists. Perhaps the only criteria for assessing which set of sentiments is more impartial is the degree to which the set is open-minded and facilitates critical inquiry. If both solutions seem to do this, then our decision as to which is better must be an individual choice in belief in James' sense. The important point is that this choice is empirically informed by empathetically bringing home each perspective to ourselves and that it is impartial insofar as we maintain the possibility that our belief might be mistaken and ought to be revised.

While I have been critical of some aspects of Smith's view here, my goal in this chapter has not been to renounce his general sentimentalist, empathy-based approach but rather to refine it based on the pragmatist approach to moral inquiry. Smith's account is helpful in providing an empirical explanation of the role of empathy and society in providing the tools for moral self-assessment. His account only encounters problems when it strays from anti-absolutism, as we saw with the problematic aspects of Smith's appeal to fixed moral rules, or when it strays from embracing fallibilism and individuality, as we saw with the problematic aspects of Smith's appeal to an ideally objective impartial spectator grounded in shared sentiments. Both of these problems stem from Smith's insistence that impartiality is constructed and codified from the bottom-up through empathetic processes. This is not the appropriate role for empathy in moral inquiry. As I have been arguing throughout this dissertation, we ought not rely on empathy to motivate impartiality, but rather ought to realize that our desire to be impartial should motivate us to engage in continuous empathetic critical assessment of our moral perspectives. This approach to impartial, empathetic moral inquiry as an empirical, fallibilistic method is aligned with the insights from pragmatism discussed in this chapter.

We can preserve what Smith gets right, that moral-self assessment and the development of moral sentiments are intimately linked to our ability to empathize with the moral sentiments of

others, while employing pragmatist points about anti-absolutism and fallibilism to address the problems of bias and relativism that face Smith's account. For the most part, Smith's account is an amenable fit with pragmatist ethics, particularly because of its empirical and social emphasis and because of the prominent role of empathy on which it relies. Thus, in combining the insights of Smith's sentimentalism with those of pragmatist ethics, we are left with an account of moral inquiry that situates empathy at its core, yet also has the tools to correct for empathy biases.

Conclusion

The goal of this dissertation has been to examine the role of empathy in moral inquiry. When the arguments of the preceding chapters are considered together, we arrive at an account of this role that can be summarized by the following claims:

First, I have argued that empathy involves both affective matching and nuanced, effortful other-oriented perspective taking, and that empathy occurs in degrees. Empathy should be distinguished from reactive, unreflective emotional contagion, and it should be distinguished from a merely cognitive, theory-theory approach to understanding other minds. Empathy involves simulating another person's experience to some degree, and this process involves appropriately contextualizing an emotional experience through effortful perspective-taking in particular cases, as well as the effortful development of a more wide-ranging, fine-grained emotional capacity.

Next, I argued that although empathy is susceptible to multiple forms of bias that are morally problematic, these biases are not insurmountable. Once we understand that our emotions are not innate, universal, and reactive, but rather are constructed according to concepts shaped by our unique experiences, we can see that the emotional capacity to empathetically engage in more wide-ranging and nuanced perspective-taking is shaped by the experiences and interactions that we choose. We can correct empathy bias.

Finally, I argued that we ought to correct empathy bias, rather than tamp down empathy, because empathy plays a valuable role in moral inquiry. It enables us to challenge our own moral assumptions from another perspective in concrete cases of disagreement (especially in cases of subtle disagreement), compassionately recognize others as individuals with moral perspectives

worth considering, incorporate the moral perspectives of others in the process of our own moral development, and realize impartiality based in fallibilistic method.

In sum, the appropriate role for empathy in moral inquiry is as a tool that allows us to analyze and value the moral perspectives of others in the pursuit of a compassionate and impartial moral life. Empathy allows us to critique moral dilemmas and moral beliefs, including our own, in a light that has incorporated empathetic insights drawn from those whose perspectives do not necessarily share our own experiential foundations nor share the biases and assumptions that may remain hidden to us if we do not make the critical effort to see morally salient actions and views as others see them.

I have defended this account of the role of empathy in moral inquiry over the course of 6 chapters.

In Chapter 1, I distinguish empathy from involuntary affective matching. The kind of empathy that is my focus requires actively engaging in other-oriented perspective taking to contextualize affective experience relative to the perspective of the target of empathy. The phenomenon I have in mind is not cognitive empathy or a theory-theory ToM; it involves some degree of simulation. Research in ToM suggests that in order to understand the emotions of others, some level of simulation will be involved. The best model may be a simulation theory model or a hybrid theory (involving both simulation and theorizing) model, but my claim is that, in either case, empathetic simulation of emotion is involved at some level in understanding the emotions of others. The focus of the dissertation is the role of this simulational, empathetic understanding of the moral emotions of others when engaging in moral inquiry.

Chapter 2 focused on empirical evidence that suggests that empathy is susceptible to problematic biases. I discussed evidence of empathy bias from social psychology and

neuroscience, not with the goal of refuting this evidence, but with the goal of motivating much of the argumentative work in the remainder of the dissertation; my aim throughout has been to argue that empathy bias is both correctable and worth correcting despite the legitimacy of this evidence. I outlined evidence of empathy bias of three kinds: (1) intergroup empathy bias, a tendency to empathize with members of ingroups over outgroups; (2) bias of scope, a bias towards empathizing with specific individuals at the expense of ignoring larger groups; (3) bias of proximity/exposure, bias towards favoring the individuals or groups to whom we are exposed via arbitrary geographic proximity or media coverage.

In Chapter 3 I defended Lisa Feldman Barrett's Conceptual Act Theory (CAT) of emotion and argued that this theory allows us to understand empathy bias as something that can be corrected with active effort. According to CAT, emotions are constructed based on emotion concepts that we develop through experience. As such, if we actively pursue more diverse experience, we will develop more wide-ranging, fine-grained emotion concepts. When we empathize, we try to construct the emotions of others, and we will be better able to do this if we actively develop our emotion concepts by seeking out experiences, communication, and engagement with art that help us achieve more potential areas of emotional overlap with those with whom we might empathize.

Section 1 of this chapter provided an in-depth consideration and defense of empirical evidence in favor of Barrett's theory. In Section 2 I argued that the conceptualization of empathy articulated in Chapter 1 can be understood in terms of CAT and defended three general strategies we can pursue to refine our emotion concepts and thus become better empathizers. Those three strategies are: 1) the embedded approach, in which we directly pursue diverse experiences such that we may share more of an experiential background with those with whom we might

empathize; 2) the communicative approach, in which we seek to communicate with those who have had experiences different from our own so as to better understand the emotions involved in those experiences; 3.) the imaginative approach, in which we engage with narrative artworks that portray emotional and moral perspectives with a depth and pace that facilitates moral and emotional reflection and development. The takeaway of Chapter 3 is that once we understand emotion in terms of CAT, we can see that empathy bias can be corrected with effort, and these three approaches provide an account of what that sort of effort involves.

My goal in Chapter 3 was to show that it is possible to correct empathy bias with effort. In Chapter 4 I turned to the question of whether we *should* make this effort to correct empathy bias rather than pursue moral inquiry that does not rely on empathy. I responded to two critics of empathy: Paul Bloom and Jesse Prinz. In doing so, I provided my own arguments about the significance of empathy as a means of critical self-reflection, as a compassionate response to other moral agents, and as a catalyst of moral development.

Bloom argues that we ought to replace empathy with what he calls "rational compassion." I argued that empathy enables us to step outside of ourselves and challenge our assumptions about which sort of moral solutions something like rational compassion calls for in the first place. Without empathetic consideration of other moral perspectives, we may become complacent and resistant to allowing our perspective on what count as rational, compassionate solutions to moral problems to develop. Empathy plays a role in the pursuit of impartiality in moral inquiry. This is most apparent in more subtle moral disputes in which differing, incompatible solutions both have legitimate claims to being compassionate. Bloom is right that empathy can be problematic in some cases, but we ought not ignore its ability to help us to engage in productive moral self-assessment in other cases. Furthermore, Bloom is right that we

ought to be motivated by compassion, but I argue that empathizing with the moral perspectives of others is compassionate. Empathy is a compassionate response to others because it is a way of recognizing the authenticity of individuals, of demonstrating that their views are significant worth caring about. We should make the effort to empathize with the moral perspective of others to demonstrate that we respect them as moral agents who can make valuable contributions to moral inquiry.

Prinz, another of empathy's strongest critics, argues that it is not empathy, but rather other moral emotions such as guilt, moral outrage, love etc. that should be cultivated in our moral lives. Prinz argues that empathy is often not and in fact should not be involved in our moral judgments. I argued that, like Bloom, Prinz is right that there are many cases in which empathy need not play a role. But I defend a role for empathy in productive, self-critical inquiry that both Bloom and Prinz neglect. I argued that the "agent empathy" of which Prinz is critical enables us to productively examine the relationship between motivation and action in our moral judgments. Prinz also claims that empathy is not a driving factor in moral development. He makes the point that it is not a psychopath's lack of empathy that leads to a lack of moral development, but rather is a more general "shallow affect." I objected to this account, arguing that the explanation of the psychopath's shallow affect involves his or her lack of empathy. Moral development is stunted in the psychopath because of an inability to empathize with others' moral emotions as directed at oneself, and this suggests the value of such empathy to the moral development of non-psychopaths.

The idea that underlies the arguments in Chapter 4 is that there is a role for empathy that critics like Bloom and Prinz overlook. The appropriate role for empathy is not to motivate impartiality, compassion, or moral growth. We take these as foundational commitments in our

moral lives, but when we do so, we see that our commitments to compassion, impartiality, and moral growth ought to motivate us to empathize in certain cases. Empathy is a tool that can be used in the service of achieving these foundational commitments. It is not opposed to them.

Following my defense of the value of empathy for moral inquiry, I further explored the relationship between empathy and impartiality. I discussed Smith's account of empathy and the impartial spectator in order to highlight some of the similarities that it shares with my own account and to highlight two significant problems that face Smith's account, problems that arise from his particular conception of impartiality and that ought to be addressed by any account that defends empathy as part of an impartial approach to moral inquiry, as my account does.

Smith's account and my account share an emphasis on the role of empathy in an empirical and social approach to impartial moral inquiry. For Smith, we ought to empathize with others to assess our own conduct from outside of our own perspective and to avoid biased assessment grounded in what he calls "self-love." Smith's account is empirical and social in that engaging in moral inquiry must involve empathizing with people that we actually encounter in the world and trying to understand how they react to real moral problems. However, I argued that the particular empirical and social nature of Smith's account of impartiality leaves it open to two difficulties: (1) Empathetic moral inquiry may lead one to construct a biased conception of impartiality, as empathy bias leads one to only consider a certain kind of perspective as relevant in the empathetic construction of Smith's idealized impartial spectator. (2) Such inquiry may lead to moral relativism, as it is not clear how the method itself is able to decide which of two competing perspectives is the more impartial without having to appeal to an impartial spectator to arbitrate and thus either begging the question or succumbing to an infinite regress.

Thus, my goal in Chapter 5 was to highlight what Smith's account gets importantly right about the role of empathy in moral inquiry, but also to highlight significant problems that this sort of account faces due to its particular conception of impartiality.

I addressed these problems in Chapter 6, drawing on insights from pragmatist philosophy to outline an empathy-based account of impartial moral inquiry that is empirical and social but avoids the issues facing Smith's account. Rather than locate impartiality in constructing a static, idealized impartial spectator, as Smith does, I defended an approach in which impartiality is characterized by the continuous application of a fallibilistic method that involves empathetic effort. I outlined how this approach draws on insights from pragmatist philosophy (though it does not rely on a defense of pragmatism in general or on pragmatist theories of truth)**,** then applied this pragmatist-influenced account of impartiality to the problems facing Smith's account, arguing that moral inquiry that is impartial in the pragmatist, fallibilistic sense can avoid the problems facing Smith's account while still realizing the benefits of empathy that I defended in previous chapters.

Smith's account runs into problems because of its reliance on general moral rules and its inability to allow for a role for individuality. For Smith, empathy is used as a means of solidifying a fixed impartial spectator constructed based on empathizing with the "shared sentiments" of others so as to limit the influence of the individual and avoid "self-love." On the other hand, a commitment to impartiality as fallibilistic method is a commitment that eschews solidifying rules and instead favors consideration of context and continuous openness to change; impartiality is not a perspective at which we arrive, but rather is a mindset that we maintain in a process of empathetic, open-minded inquiry. Importantly, this impartial method leaves room for the individual in moral action and does not necessarily defer to shared sentiments. One

empathetically considers the views of others as evidence relevant to the moral problem at hand, but ultimately the individual is the locus of moral action and judgment; action and judgment is dictated by individual conviction. But the key is that if we are impartial, our conviction remains open to reconsideration and potential revision based on the perspective of others. Thus, the assessment of a moral action or judgment is not entirely reliant on "shared sentiments," but it is still a critical assessment in which the sentiments of others have been thoroughly, legitimately, empathetically considered as evidence relevant to one's own individual actions and convictions.

I want to conclude by briefly returning to what I consider to be two of the most pressing questions that my account of the role of empathy in moral inquiry has had to address. These are questions that I have addressed at various points throughout the dissertation in relation to specific arguments, but which I think are worth discussing in a more holistic fashion here now that I have presented my account in its entirety. I believe that doing so will both solidify my over-arching claims about the role of empathy in moral inquiry and suggest avenues for future development that may be pursued with this role in mind.

First, there is the normative question of the appropriate limits of empathetic effort. If we agree that there is value in effortful empathetic engagement with other moral perspectives, we still need to ask how far to pursue such effort. Where do we draw lines in terms of trying to empathize with differing moral perspectives? Second, there is the descriptive question of the limits of our empathetic abilities. To what extent are we truly capable of empathizing with moral perspectives that differ from our own?

I addressed the normative question in my defense of what I called the horizons of compassion in Chapter 4. I argued that we ought to empathize with perspectives that have a legitimate claim to being compassionate, and that solutions ought to be considered

compassionate for reasons, rather than due to mere stipulation. Of course, this raises the following crucial question for moral inquiry: what are the relevant reasons that qualify a proposed solution to a moral problem as compassionate?

Answering this question, particularly in terms of specific problems in applied areas such as environmental ethics, humanitarian aid, and animal ethics, ought to be a goal of moral inquiry, but my goal here has not been to focus on providing such answers; rather, it has been to highlight the role of empathy in a method of critiquing solutions that fall within the horizons of compassion, solutions that ought to count as live options for us to consider. Determining the best solution to a given moral problem requires not just assessing which consequences might occur given a proposed solution and assessing the probability of the occurrence of those consequences. It also requires the valuation of those consequences. As I have argued throughout, the role of empathy in moral inquiry is particularly apparent in disputes in which there is general agreement about the matters of fact regarding proposed solutions, but disagreement regarding how to value those facts. In empathizing with the moral perspectives of others, we allow ourselves to engage in a process of valuation of moral solutions that incorporates emotions and experiences that differ from those on which our own initial valuations are based, and thus to assess different schemas of valuation themselves rather than merely assessing different consequences through the filter of our own schema.

It has not been my aim here to defend a normative theory about which specific solutions this empathetic consideration of different valuations should lead us to. My aim has been to defend the value of the method of empathetic consideration itself. Indeed, I think it is a feature of the sort of pragmatist, fallibilistic approach to inquiry that I have defended that we ought not try to determine a solution in advance of the process of inquiry. When engaging in moral inquiry, we

ought to maintain an open mind to valuations that differ from our own, and to allow our empathetic capacity to assist us in doing so. While an open mind does not prevent us from grounding our inquiry in basic features of compassion such as equity and avoidance of harm, these general considerations will always need to be cashed out and weighted in terms of the particular moral problem at hand. We will still need to ask what it means to realize equity or harm avoidance in terms of the problem we are addressing, and we still need to find the appropriate balance of these considerations as they relate to that particular problem. Drawing on a foundation of basic features of compassion is helpful in allowing us to realize that a solution cannot merely be stipulated as compassionate, but we still need to consider the particulars of the problem at hand when considering which solutions best meet our basic criteria.

Deciding which solutions fall within the horizons of compassion is no doubt an important part of moral inquiry, but this process need not rely on empathy. What I have argued is that, once we have, based on relevant reasons, decided which views fall within the horizons of compassion in a particular situation, empathy enables us to better understand different valuations of the factors that we have agreed are in fact relevant to realizing compassionate solutions. Take, for example, disagreement about how to mitigate climate change. We may agree that our solution should seek to prevent harm and roughly agree on what the empirical consequences of various solutions would be, yet we may still disagree about the value we should place on considerations such as animal and plant life, future generations, or harms to human beings due to certain economic adjustments. In other words, agreement on the empirical effects of a solution will not necessarily yield agreement on the value of that solution. Here is a situation in which an impasse regarding valuation can be helped by effortfully trying to empathize with a perspective that, say, feels guilt regarding the destruction of natural habitats, or that feels that failing to value a certain

economic opportunity or quality of life for human beings is cold-hearted and harmful. When engaging in moral inquiry regarding this problem, we ought to make an effort to empathetically imagine the emotions and experiences that underly other perspectives and see whether our own perspectives look different in light of doing so. We ought to take seriously the question: how could someone arrive at that alternative valuation? This does not mean that we need to agree with the alternative valuation in question, it merely means that we ought to make the effort to empathetically consider the perspective behind it, as doing so allows us to understand our own view in a more critical, distanced manner. We may maintain our valuation despite this process, or slightly alter our valuation, or alter our valuation significantly, but the important point is that we have engaged in a fallibilistic, self-critical process of inquiry by making the effort to empathize, and we ought to have more confidence in the impartiality of our valuation as a result.

I should again stress what empathy is not doing in this process. It is not telling us what consequences we ought to consider. It is telling us about different ways of valuing those consequences. Importantly, empathy is also not necessarily leading us to agree with these different valuations. It is leading us to consider our own valuations in relation to legitimate alternatives such that we open the possibility of recognizing biases and assumptions of which we were previously unaware.

This leads to the descriptive question: how capable are we of empathizing with the moral perspectives of others? I addressed this question in my defense of the role of experience in developing emotion concepts as defined by CAT, and in my emphasis on the value of empathy as a phenomenon that occurs in degrees. According to CAT, our emotion concepts are fundamentally defined by our unique experiential backgrounds. Emotions are not universal. I have stressed the importance of this idea in terms of the malleability of emotions, the potential to

shape our emotional lives and thus our empathetic abilities via the experiences that we consciously choose. However, there is another side to this idea that a critic of empathy may emphasize, namely that if our emotion concepts are defined by unique experiences, then there is a sense in which no two people will be able to experience the exact same emotion, given that no two people will share the exact same experiential background.

This implication is only problematic for a defense of empathy that relies on a conception of empathy that involves complete emotional identification between empathizer and the target of empathy. I do not think such a conception is realistic and it is not the phenomenon that I have defended here. If emotions were in fact simple and universal, then perhaps we could achieve such empathy in the same way that we can achieve matching of basic affective valence, but I have argued, following Barrett, that the science of emotion does not support such an account of basic universal emotions. Furthermore, I think this lack of basic universal emotions is encouraging for the benefits of empathy in moral inquiry; to understand emotions as complex concepts based in unique experiences is to understand that it is worth making an effort to empathize with the emotions of others so as to understand who they are as unique individuals. Empathizing with simple, universal emotions, while certainly easier than empathizing with complex emotions, would tell us little about unique individual perspectives, given that *ex hypothesi* we already share the universal emotion response with which we would empathize. While CAT tells us that we may not be able to achieve complete overlap with the emotion concepts of others, it does not tell us that there is no potential for degrees of overlap. We are capable of developing fine-grained emotion concepts that allow us to achieve higher degrees of overlap with others. These degrees of overlap between complex, unique emotional perspectives are more informative than a complete overlap of simple emotions or of mere affective valence.

We may only be able to achieve degrees of empathy with other perspectives, but these degrees of empathetic connection are valuable in the sense that they tell us about a degree of the other's unique perspective and may encourage us to seek further connection that continues to refine our own emotional palette and our moral understanding. And a realization of the limits of one's empathy is valuable. When we fail to empathize with some aspect of another's perspective despite a legitimate effort at perspective taking, this process tells us about the emotional and experiential barriers that define our own perspectives.

Advocating for the value of empathy in moral inquiry is not to advocate for the unrestrained empathy for all moral views, nor is it to advocate for the complete empathetic understanding of any one particular moral view. It is to defend the value of an open-minded effort to empathetically imagine some degree of the perspectives of those whose moral views differ from our own; that is, to not allow moral differences to block moral inquiry.

A theme that has run throughout my defense of empathy is the relationship between the individual and society in moral inquiry. I have argued, particularly in Chapter 4, that empathetic engagement with society can prevent the unchecked reification of one's personal convictions as absolute moral truths. But I have also argued, particularly in Chapter 6, that one needs to rely on personal conviction to prevent the potential reification of societal biases as absolute moral truths. There is an obvious tension between these two claims and understanding the value of empathy in degrees is critical in trying to resolve this tension. The limitations of our empathetic abilities allow us to achieve a constructive balance of the self and society in fallibilistic moral inquiry. Degrees of empathy for the perspectives of others help us understand different valuations regarding moral problems without necessarily endorsing those valuations. Empathetic experience may change our own valuations, or it may not, and ultimately the individual, and not mere

adherence to societal convention, is the locus of this change. Empathetic experience enables an individual to bring home other perspectives to himself or herself in a manner such that other perspectives are legitimately considered in moral inquiry, but empathetic experience is not a complete overcoming of the individual's own perspective and agency, it is an experience of degrees of understanding of the perspectives of others. Ultimately these degrees of understanding ought to play a role in the consideration of the problem at hand, but it is not the case that one merely gives up one's moral identity and agency in empathizing with the moral perspectives of others, and this is precisely because empathy is a phenomenon that occurs in degrees.

I have defended an approach to moral inquiry that utilizes empathy as a means of careful consideration of the perspectives of other moral agents, and that draws on empathetic effort in the process of self-critical moral reflection. Empathetic moral inquiry is a fallibilistic process focused on the particulars of moral problems and respectful of the nuances of different solutions to those problems. It allows us to challenge our own values by inhabiting the perspectives of others and to compassionately recognize that the perspectives of others are in fact worth making the effort to inhabit. I think that Adam Smith is right that society is a mirror through which we can view the morality of our own conduct and views, and I think that the role of empathy in moral inquiry is to seek a clarity of reflection.

References

Addams, J. (2005). *Democracy and social ethics*. Project Guttenberg. (Original work published 1902).

Anderson, E. (2018). Uses of value judgments in science: A general argument, with lessons from a case study of feminist research on divorce. *Critical Realism, History, and Philosophy in the Social Sciences*, *34*, 47–71. https://doi.org/10.1108/S0198-871920180000034003

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., Anderson, A., Lee, G. P., & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, *37*(10), 1111–1117. https://doi.org/10.1016/s0028-3932(99)00039-1

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*(1), 21–62. https://doi.org/10.1177/1534582302001001003

Alexander, M. (2012). *The new Jim Crow: Mass incarceration in the age of color blindness*. The New Press.

Allport, G. (1954). *The nature of prejudice*. Addison Wesley.

Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review*, *106*(3), 529–550. https://doi.org/10.1037/0033-295x.106.3.529

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*(10), 1531–1542. https://doi.org/10.1177/0956797615594620

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions.

Trends in Cognitive Sciences, *11*(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310

Barrett, L. F. (2005). Feeling is perceiving: Core affect and conceptualization in the experience of Emotion. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 255–284). The Guilford Press.

Barrett, L. F. (2011). Constructing emotion. *Psychological Topics, 20*(3), 359–380.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science, 20*(5), 286–290. https://doi.org/10.1177/0963721411422522

Barrett, L. F. (2015). Ten common misconceptions about psychological construction theories of emotion. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 45–79). The Guilford Press.

Barrett, L. F., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). The conceptual act theory: A roadmap. In L. F. Barrett & J. A. Russell (Eds.), *The psychological construction of emotion* (pp. 83–110). The Guilford Press.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature reviews. Neuroscience*, *16*(7), 419–429. https://doi.org/10.1038/nrn3950

Barrett, L. F. (2018). *How emotions are made*: *The secret life of the brain*. Mariner Books.

Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in cognitive sciences*, *7*(2), 84–91. https://doi.org/10.1016/s1364-6613(02)00029-3

Barsalou L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1521), 1281–1289. https://doi.org/10.1098/rstb.2008.0319

Batson, C. D., Klein, T. R., Highberger, L., & Shaw, L. L. (1995). Immorality from empathy-induced altruism: When compassion and justice conflict. *Journal of Personality and Social Psychology, 68*(6), 1042–1054. https://doi.org/10.1037/0022-3514.68.6.1042

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., Klein, T. R., & Highberger, L. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology, 72*(1), 105–118. https://doi.org/10.1037/0022-3514.72.1.105

Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–15). MIT Press. https://doi.org/10.7551/mitpress/9780262012973.003.0002

Batson, C. D., & Ahmad, N. Y. (2009). Using empathy to improve intergroup attitudes and relations. *Social Issues and Policy Review, 3*(1), 141–177. https://doi.org/10.1111/j.1751-2409.2009.01013.x

Batson, C.D. (2011). *Altruism in humans*. Oxford University Press.

Beard, M. (2014). *Laughter in ancient Rome: On joking, tickling, and cracking up*. University of California Press.

Becker, B., Mihov, Y., Scheele, D., Kendrick, K. M., Feinstein, J. S., Matusch, A., Aydin, M., Reich, H., Urbach, H., Oros-Peusquens, A. M., Shah, N. J., Kunz, W. S., Schlaepfer, T. E., Zilles, K., Maier, W., & Hurlemann, R. (2012). Fear processing and social networking

in the absence of a functional amygdala. *Biological psychiatry*, *72*(1), 70–77. https://doi.org/10.1016/j.biopsych.2011.11.024

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state. A functional MRI study. *Journal of Cognitive Neuroscience*, *11*(1), 80–95. https://doi.org/10.1162/089892999563265

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767-2796. http://dx.doi.org/10.1093/cercor/bhp055

Black, J., & Barnes, J. L. (2015). Fiction and social cognition: The effect of viewing award-winning television dramas on theory of mind. *Psychology of Aesthetics, Creativity, and t he Arts, 9*(4), 423–429. https://doi.org/10.1037/aca0000031

Bloom, P. (2016). *Against empathy: The case for rational compassion*. HarperCollins.

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *NeuroImage*, *25*(1), 312–319. https://doi.org/10.1016/j.neuroimage.2004.11.043

Broadie, A. (2006). Sympathy and the impartial spectator. In K. Haakonssen (Ed.) *The Cambridge companion to Adam Smith*. Cambridge University Press.

Brown, L. M., Bradley, M. M., & Lang, P. J. (2006). Affective reactions to pictures of ingroup and outgroup members. *Biological psychology*, *71*(3), 303–311. https://doi.org/10.1016/j.biopsycho.2005.06.003

Bruneau, E. G., & Saxe, R. (2012). The power of being heard: The benefits of "perspective-

giving" in the context of intergroup conflict. *Journal of Experimental Social Psychology, 48*(4), 855–866. https://doi.org/10.1016/j.jesp.2012.02.017

Calder, A. J., Keane, J., Cole, J., Campbell, R., & Young, A. W. (2000). Facial expression recognition by people with mobius syndrome. *Cognitive neuropsychology*, *17*(1), 73–87. https://doi.org/10.1080/026432900380490

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy, 97*(5), 1232-1254.

Carroll, N. (2001). *Beyond aesthetics: Philosophical essays*. Cambridge University Press.

Carroll, N. (2011). On some affective relations between audiences and the characters in popular fictions. In A. Coplan & P. Goldie (Eds.), *Empathy: philosophical and psychological* perspectives. (pp. 162–184). Oxford University Press.

Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011a). Us and them: Intergroup failures of empathy. *Current Directions in Psychological Science: A Journal of the American Psychological Society*, *20*(3), 149–153. https://doi.org/10.1177/0963721411408713

Cikara, M., & Fiske, S. T. (2011). Bounded empathy: neural responses to outgroup targets' (mis)fortunes. *Journal of Cognitive Neuroscience*, *23*(12), 3791–3803. https://doi.org/10.1162/jocn_a_00069

Cikara, M., Botvinick, M. M., & Fiske, S. T. (2011c). Us versus them: social identity shapes neural responses to intergroup competition and harm. *Psychological Science*, *22*(3), 306–313. https://doi.org/10.1177/0956797610397667

Cikara, M., Bruneau, E., Van Bavel, J. J., & Saxe, R. (2014). Their pain gives us pleasure: How intergroup dynamics shape empathic failures and counter-empathic responses. *Journal of experimental social psychology*, *55*, 110–125. https://doi.org/10.1016/j.jesp.2014.06.007

Citrin, J., & Sides, J. (2008). Immigration and the imagined community in Europe and the United States. *Political Studies*, *56*(1), 33–56. https://doi.org/10.1111/j.1467-9248.2007.00716.x

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and brain sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Coates, T.-N. (2014). The case for reparations. *The Atlantic Monthly, 313*(5), 54–71.

Combs, D. J., Powell, C. A., Schurtz, D. R., & Smith, R. H. (2009). Politics, schadenfreude, and ingroup identification: The sometimes happy thing about a poor economy and death. *Journal of Experimental Social Psychology*, *45*(4), 635–646. https://doi.org/10.1016/j.jesp.2009.02.009

Coplan, Amy. (2011a). Understanding empathy: Its features and effects. In A. Coplan & P. Goldie (Eds.), *Empathy: philosophical and psychological* perspectives. (pp. 3–18). Oxford University Press.

Coplan, Amy. (2011b). Will the real empathy please stand up? A case for a narrow conceptualization. *Southern Journal of Philosophy*, *49*, 40–65. https://doi.org/10.1111/j.2041-6962.2011.00056.x

Cuddy, A. J. C., Rock, M. S., & Norton, M. I. (2007). Aid in the aftermath of Hurricane Katrina: inferences of secondary emotions and intergroup helping. *Group Processes & Intergroup Relations, 10*(1), 107–118. https://doi.org/10.1177/1368430207071344

Currie, G. (1995). The moral psychology of fiction. *Australasian Journal of Philosophy*, *73*(2), 250-259, https://doi.org/10.1080/00048409512346581

Dalgleish, T. & Power, M. J. (Eds.). (1999). *Handbook of cognition and emotion*. Wiley.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions.

Proceedings of the National Academy of Sciences of the United States of America, *108*(17), 6889–6892. https://doi.org/10.1073/pnas.1018033108

Dashiell, J. F. (1927). A new method of measuring reactions to facial expression of emotion. *Psychological Bulletin*, 24, 174-175.

Davies, S. (2011). Infectious music: music-listener emotional contagion. In A. Coplan & P. Goldie (Eds.), *Empathy: philosophical and psychological* perspectives. (pp. 3–18). Oxford University Press.

Darwin, Charles (2005). *The expression of the emotions in man and animals*. Digireads. (Original work published 1872).

Dewey, J. (1993). The ethics of democracy. In D. Morris & I. Shapiro (Eds.) *John Dewey: The political writings*. Hackett. (Original work published 1888)

Dewey, J. (1910). *How we think.* In J. Boydston (Ed.) *The middle works: 1899-1924*. Southern Illinois University Press. (Original work published 1910).

Dewey, J. (1916a). *Democracy and education*. In J. Boydston (Ed.) *The middle works: 1899-1924*. Southern Illinois University Press.

Dewey, J. (1916b). *Essays in experimental logic*. Dover.

Dewey, J. & Tufts, J. H. (1945). *Ethics*. Holt.

Dewey, J. (1939). Theory of valuation. In O. Neurath (Ed.) *International encyclopedia of unified science*. The University of Chicago Press.

Dweck, C. S. (2006). *Mindset: the psychology of success*. Random House.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129. https://doi.org/10.1037/h0030377

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of

emotion. *Science, 164*(3875), 86–88. https://doi.org/10.1126/science.164.3875.86

Enos, R. D. (2014). Causal effect of intergroup contact on exclusionary attitudes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(10), 3699–3704. https://doi.org/10.1073/pnas.1317670111

Fesmire, S. (2003). *John Dewey and moral imagination: Pragmatism in ethics*. Indiana University Press.

Fleischacker, S. (2011). True to ourselves? Adam Smith on self-deceit. In F. Forman-Barzilai (Ed.) *Adam Smith Review*, (Volume 6). Routledge.

Friston K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Cultural relativity in perceiving emotion from vocalizations. *Psychological science*, *25*(4), 911–920. https://doi.org/10.1177/0956797613517239

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014b). Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion (Washington, D.C.)*, *14*(2), 251–262. https://doi.org/10.1037/a0036052

Goldenberg, A., Cohen-Chen, S., Goyer, J. P., Dweck, C. S., Gross, J. J., & Halperin, E. (2018). Testing the impact and durability of a group malleability intervention in the context of the Israeli–Palestinian conflict. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 115*(4), 696–701. https://doi.org/10.1073/pnas.1706800115

Goldman, A. (2006). *Simulating minds*. Oxford University Press.

Gordon, R. (1995). Sympathy, simulation, and the impartial spectator. *Ethics, 105*(4), 727-742.

Gray, J. R., Schaefer, A., Braver, T. S., & Most, S. B. (2005). Affect and the Resolution of

Cognitive Control Dilemmas. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 67–94). The Guilford Press.

Hatfield, E., Cacioppo, J.T., & Rapson, R.L. (1994). *Emotional contagion*. Cambridge University Press.

Hatfield, E., Rapson, R. L., & Le, Y.-C. L. (2009). Emotional contagion and empathy. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 19–30). MIT Press. https://doi.org/10.7551/mitpress/9780262012973.003.0003

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, *68*(1), 149–160. https://doi.org/10.1016/j.neuron.2010.09.003

Hainmuller, John and Hopkins, Daniel J. (2014) "Public Attitudes Toward Immigration" *Annual Review of Political Science,* 17, 225-249.

Hong, Y.-y., Chiu, C.-y., Dweck, C. S., Lin, D. M.-S., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77*(3), 588–599. https://doi.org/10.1037/0022-3514.77.3.588

Hume, D. (2000). *A treatise of human nature*, (D. F. Norton and M.J. Norton, Eds.), Oxford University Press. (Original work published 1739)

Izard, C. E. (1971). *The face of emotion*. Appleton-Century-Crofts.

Izard C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychological bulletin*, *115*(2), 288–299. https://doi.org/10.1037/0033-2909.115.2.288

James, W. (1891a). The moral philosopher and the moral life. In *The will to believe and other essays in popular philosophy and human immortality*. Dover.

James, W. (1891b). The will to believe. In *The will to believe and other essays in popular philosophy and human immortality*. Dover.

James, W. (1907). *Pragmatism*. Hackett.

Jeske, D. (2018). *The evil within: Why we need moral philosophy*. Oxford University Press.

Johnson, J. D., Simmons, C. H., Jordan, A., MacLean, L., Taddei, J., Thomas, D., Dovidio, J. F., & Reed, W. (2002). Rodney King and O. J. revisited: The impact of race and defendant empathy induction on judicial decisions. *Journal of Applied Social Psychology, 32*(6), 1208–1223. https://doi.org/10.1111/j.1559-1816.2002.tb01432.x

Johnston, B. M., & Glasford, D. E. (2018). Intergroup contact and helping: How quality contact and empathy shape outgroup helping. *Group Processes & Intergroup Relations, 21*(8), 1185–1201. https://doi.org/10.1177/1368430217711770

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54–86. https://doi.org/10.1017/bpp.2016.2

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-292.

Keen, S. (2007) *Empathy and the Novel*. Oxford University Press.

Kidder, W. (2021). The aesthetic and cognitive value of empathy for rough heroes. *Journal of Value Inquiry*, *19*(1).

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *NeuroImage*, *42*(2), 998–1031. https://doi.org/10.1016/j.neuroimage.2008.03.059

Kogut, T. & Ritov, I. (2005). The 'identified victim' effect: An identified Group, or just a single individual?" *Journal of Behavioral Decision Making*, *18*(3), 157–167.

Kogut, T., Ritov, I., Rubaltelli, E., & Liberman, N. (2018). How far is the suffering? The role of psychological distance and victims' identifiability in donation decisions. *Judgment & Decision Making*, *13*(5), 458–466.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage*, *24*(3), 771–779. https://doi.org/10.1016/j.neuroimage.2004.09.006

Larsen, J. T., Berntson, G. G., Poehlmann, K. M., Ito, T. A., & Cacioppo, J. T. (2008). The psychophysiology of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 180–195). The Guilford Press.

Lawrence, A. D., Calder, A. J., McGowan, S. W., & Grasby, P. M. (2002). Selective disruption of the recognition of facial expressions of anger. *Neuroreport*, *13*(6), 881–884. https://doi.org/10.1097/00001756-200205070-00029

Lerner, J. S., Small, D. A., & Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological science*, *15*(5), 337–341. https://doi.org/10.1111/j.0956-7976.2004.00679.x

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27*(4), 363–384. https://doi.org/10.1111/j.1469-8986.1990.tb02330.x

Levenson, R. W., Ekman, P., Heider, K., & Friesen, W. V. (1992). Emotion and autonomic nervous system activity in the Minangkabau of west Sumatra. *Journal of personality and social psychology*, *62*(6), 972–988. https://doi.org/10.1037//0022-3514.62.6.972

Leyens, J.-P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review, 4*(2), 186–197. https://doi.org/10.1207/S15327957PSPR0402_06

Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion (Washington, D.C.)*, *6*(1), 125–138. https://doi.org/10.1037/1528-3542.6.1.125

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *The Behavioral and brain sciences*, *35*(3), 121–143. https://doi.org/10.1017/S0140525X11000446

Lovejoy, A. (1908). The thirteen pragmatisms. *The Journal of Philosophy Psychology and Scientific Methods*, *5*(1), 29–59.

Lutz, C. (1983). Parental goals, ethnopsychology, and the development of emotional meaning. *Ethos, 11*(4), 246–262. https://doi.org/10.1525/eth.1983.11.4.02a00040

Masten, C. L., Gillen-O'Neel, C., & Brown, C. S. (2010). Children's intergroup empathic processing: the roles of novel ingroup identification, situational distress, and social anxiety. *Journal of Experimental Child Psychology*, *106*(2-3), 115–128. https://doi.org/10.1016/j.jecp.2010.01.002

Mathur, V. A., Harada, T., Lipke, T., & Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *NeuroImage*, *51*(4), 1468–1475. https://doi.org/10.1016/j.neuroimage.2010.03.025

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280–293. https://doi.org/10.1016/j.tics.2018.02.001

Morton, A. (2011). Empathy for the devil. In A. Coplan & P. Goldie (Eds.), *Empathy: philosophical and psychological* perspectives. (pp. 318–330). Oxford University Press.

Newton, E. (1990). Overconfidence in the communication of intent: Heard and unheard melodies. Unpublished doctoral dissertation. Stanford University.

Nickerson, R. S., Butler, S. F., & Carlin, M. (2009). Empathy and knowledge projection. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 43–56). MIT Press. https://doi.org/10.7551/mitpress/9780262012973.003.0005

Nussbaum, M. (1983). Flawed crystals: James's *The Golden Bowl* and literature as moral philosophy. *New Literary History, (15)*4, 25-50.

Nussbaum, M. (1985). Finely aware and richly responsible: Moral attention and the moral task of literature. *The Journal of Philosophy, 82*(10), 516-529.

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding v1? *Neural Computation*, *17*(8), 1665–1699. https://doi.org/10.1162/0899766054026639

Peirce, C. S. (1877). The fixation of belief. *Popular Science Monthly*, 12, 1-15.

Peirce, C. S. (1898). The first rule of logic. In Peirce Edition Project (Ed.) *The essential Peirce: Selected philosophical writings*. Indiana University Press.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology, 38*(6), 922–934. https://doi.org/10.1002/ejsp.504

Phelps, E. A. (2006). Emotion and cognition: insights from studies of the human amygdala.

Annual review of psychology, *57*, 27–53. https://doi.org/10.1146/annurev.psych.56.091103.070234

Porat, R., Halperin, E., & Tamir, M. (2016). What we want is what we get: Group-based emotional preferences and conflict resolution. *Journal of personality and social psychology*, *110*(2), 167–190. https://doi.org/10.1037/pspa0000043

Prinz, J. J. (2008). *The emotional construction of morals*. Oxford University Press.

Prinz, J. J. (2011a). Is empathy necessary for morality? In A. Coplan & P. Goldie (Eds.), *Empathy: philosophical and psychological* perspectives. (pp. 211–229). Oxford University Press.

Prinz, J. J. (2011b). Against empathy. *The Southern Journal of Philosophy, 49*(1), 214–233.

Putnam, H. (1990). How not to solve ethical problems. In J. Conant (ed.) *Realism with a human face*. Harvard University Press.

Putnam, H. (1992). Existential humanism. In *Recovering philosophy*. Harvard University Press.

Putnam, H. (2002). *The collapse of the fact/value dichotomy and other essays*. Harvard University Press.

Putnam, R. A. (2009). Democracy and value inquiry. In J. R. Shook & J. Margolis (Eds.) *A companion to pragmatism*. Blackwell.

Quine, W.V. (1953). Two dogmas of empiricism. In Y. Balashov & A.Rosenberg (Eds.) *Philosophy of science: contemporary readings*. Routledge.

Rick, J. (2007). Hume's and Smith's partial sympathies and impartial stances, *Journal of Scottish Philosophy* 5.2 (October 2007) 135-158.

Rifkin, J. (2009). *The empathic civilization: The race to global consciousness in a world of crisis*. Penguin Books.

Raichle, M. E. (2010). Two views of brain function. *Trends in cognitive sciences*, *14*(4), 180–190. https://doi.org/10.1016/j.tics.2010.01.008

Russell J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin*, *115*(1), 102–141. https://doi.org/10.1037/0033-2909.115.1.102

Saarela, M. V., Hlushchuk, Y., Williams, A. C., Schürmann, M., Kalso, E., & Hari, R. (2007). The compassionate brain: humans detect intensity of pain from another's face. *Cerebral cortex (New York, N.Y. : 1991)*, *17*(1), 230–237. https://doi.org/10.1093/cercor/bhj141

Schelling, T. C. (1968). The life you save may be your own. In S. B. Chase (Ed.) *Problems in public expenditure analysis* (pp.127–162). Brookings Institution.

Selvanathan, H. P., Techakesari, P., Tropp, L. R., & Barlow, F. K. (2018). Whites for racial justice: How contact with Black Americans predicts support for collective action among White Americans. *Group Processes & Intergroup Relations, 21*(6), 893–912. https://doi.org/10.1177/1368430217690908

Serpell, N. (2019). The banality of empathy. *The New York Review of Books*.

Shipp, S., Adams, R. A., & Friston, K. J. (2013). Reflections on agranular architecture: predictive coding in the motor cortex. *Trends in neurosciences*, *36*(12), 706–716. https://doi.org/10.1016/j.tins.2013.09.004

Schumann, K., Zaki, J., & Dweck, C. S. (2014). Addressing the empathy deficit: beliefs about the malleability of empathy predict effortful responses when empathy is challenging. *Journal of personality and social psychology*, *107*(3), 475–493. https://doi.org/10.1037/a0036738

Simas, E., Clifford, S., & Kirkland, J. (2020). How empathic concern fuels political polarization.

American Political Science Review, 114(1), 258–269.

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., Quigley, K. S., & Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological bulletin*, *144*(4), 343–393. https://doi.org/10.1037/bul0000128

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science (New York, N.Y.)*, *303*(5661), 1157–1162. https://doi.org/10.1126/science.1093535

Slote, M. (2010a). *Moral sentimentalism*. Oxford University Press.

Slote, M. (2010b). The mandate of empathy. *Dao* 9(3), 303–307.

Smith, A. (1982). *The theory of moral sentiments*, (D.D. Raphael and A.L. Macfie, Eds.), Liberty Fund. (Original work published 1759).

Sprengelmeyer, R., Young, A., Schroeder, U., Grossenbacher, P., Federlein, J., Büttner, T., & Przuntek, H. (1999). Knowing No Fear. *Proceedings: Biological Sciences, 266*(1437), 2451-2456.

Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable neural systems support retrieval of how and why action knowledge. *Psychological science*, *21*(11), 1593–1598. https://doi.org/10.1177/0956797610386618

Tam, T., Hewstone, M., Cairns, E., Tausch, N., Maio, G., & Kenworthy, J. (2007). The impact of intergroup emotions on forgiveness in Northern Ireland. *Group Processes & Intergroup Relations, 10*(1), 119–136. https://doi.org/10.1177/1368430207071345

Tassinary, L., & Cacioppo, J. (1992). Unobservable facial actions and emotion. *Psychological Science, 3*(1), 28-33.

Taylor, C. (1991). *The ethics of authenticity*. Harvard University Press.

Tomkins, S. S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills, 18*(1), 119–158. https://doi.org/10.2466/pms.1964.18.1.119

United Nations Refugee Agency. (2021). *Syria emergency*. https://www.unhcr.org/en-us/syria-emergency.html

Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality & Social Psychology Bulletin*, *29*(9), 1159–1168. https://doi.org/10.1177/0146167203254597

Vezzali, L., Stathi, S., Giovannini, D., Capozza, D., & Trifiletti, E. (2015). The greatest magic of Harry Potter: Reducing prejudice: Harry Potter and attitudes toward stigmatized groups. *Journal of Applied Social Psychology*, *45*(2), 105–121. https://doi.org/10.1111/jasp.12279

de Waal, F. (2009). *The age of empathy*. Three Rivers Press.

Williams, B. & Smart, J.J.C. (1973). *Utilitarianism: For and against*. Cambridge University Press.

Williams, B. (1981). *Moral luck*. Cambridge University Press.

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia*, *49*(5), 1105–1127. https://doi.org/10.1016/j.neuropsychologia.2010.12.032

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Situating emotional experience. *Frontiers in human neuroscience*, *7*, 764. https://doi.org/10.3389/fnhum.2013.00764

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2015). Variety in emotional life: within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks. *Social cognitive and affective neuroscience*, *10*(1), 62–71. https://doi.org/10.1093/scan/nsu037

Wittgenstein, L. (1953). *Philosophical investigations*. Wiley-Blackwell

World Health Organization. (2020). *Global status report on preventing violence against children 2020*. https://www.who.int/publications/i/item/9789240004191

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(26), 8525–8529. https://doi.org/10.1523/JNEUROSCI.2418-09.2009