

# Introduction

Pushing my way out of a crowded subway car, the doors closed around my neck. This was my fault—I was trying to wedge through while the doors were closing. New York subway doors normally reopen when they hit an obstacle, but not this time. My head was sticking out of the car while the rest of me thrashed around inside. None of my thrashing made any difference. I had a vision of the train speeding away with my head bobbing outside like a jousting target, and I started to panic. Some nearby people just stared, but several jumped in to help. Two of them tried to pull open the doors, while another yelled to get the conductor’s attention. A few seconds later, the doors opened, I escaped, and, in typical Manhattan fashion, everyone went on their way.

Before I get to the philosophical point of that story, here is another. In February of 2015, a photo of a dress went viral online. Some people saw the dress as blue and black, while others saw it as white and gold. Almost nobody was able to see it both ways. The philosopher John Morrison was invited to talk about the phenomenon on Alex Wagner’s “Now” show. Wagner suggested that people became fascinated with the dress because the divergent reactions made them question their grip on reality. Morrison agreed, and offered an analogy. Finding out about these divergent color experiences, Morrison said, is like encountering someone with radically different moral views for the first time. These experiences challenge our assumption that we all uniformly grasp a single, objective reality.

In the first story, there were divergent reactions as well. Some people felt moved to try to help me while others did not. Most of those who were not moved were probably just manifesting the bystander effect, but there might have been a sadist in the mix. The sadist might have looked on with pleasure and thought, “Ha! Look at those dupes trying to help that idiot.” According to one common philosophical view, the sadist’s pleased reaction would have got reality right no more or less than the helpful strangers’ reactions did. This is similar to a conclusion many people have drawn about color, namely, that no color experience gets reality more right than any other (perhaps because colors are not fully objective properties)<sup>1</sup>.

My aim in this book is to argue against this common view about sadistic vs. helpful reactions. The latter sort of reactions, I claim, manifest a form of compassion that gets certain parts of reality right in a way that no sadistic reactions do. The *parts of reality*

<sup>1</sup> Those of us who “wrongly” saw the dress as white and gold might find some consolation in such a view.

## 2 INTRODUCTION

that compassionate reactions get right are others' feelings and our relations to them. The sense in which they *get things right* is that they match others' feelings. In the above example, I desperately wanted to get out of that situation, and the compassionate strangers formed a similar desire in response. These mental states matched because they resembled each other: both my want and the strangers' desires had the same aim, namely, getting me out of the situation. No such match holds for a sadistic reaction: the sadist felt no desire for me to get out of the situation. For that reason, I hold, a sadist would have been wrong to think the compassionate people were mere dupes (though he might have been right about the idiot). By contrast, I do not think any similar argument shows that one type of color experience of the dress gets reality right. Some types of color experiences of the dress—in particular, the black and blue ones—might be more common or reliable among humans, but that does not show that those experiences match the dress itself. Reality might set no particular standard for how we should experience colors, but it does, I believe, set a standard for how we should react to others' suffering.

One way to gloss the central claim of this book is to say that *compassion is the perception of others' feelings* and that *sadistic pleasure and indifference to others' feelings are perceptual lacks*. Insofar as perception is about latching onto truth, another way to gloss the claim is to say that *cruel and indifferent people cannot face the truth*. This, I believe, has implications for morality: it answers the question of why we should be moral and, together with some further, fairly modest assumptions, provides the keystone for a grounding for morality that secures it against the threats of egoism, relativism, and other rationalizations people use for cruelty and apathy. In the terms of contemporary metaethics,<sup>2</sup> I hold that this fact about compassion's reality-matching can be used to provide a naturalistic but uncompromising form of moral realism, without requiring any metaphysical or semantic trickery. Though my way of articulating and defending this view about compassion and morality is new, the basic thought is not. As I discuss in Chapter 1, Arthur Schopenhauer endorsed it. More recently, Nel Noddings expressed the thought (or one close to it) in terms of care:

Apprehending the other's reality, feeling what he feels as nearly as possible, is the essential part of caring from the view of the one-caring. For if I take on the other's reality as possibility and begin to feel its reality, I feel, also, that I must act accordingly; that is, I am impelled to act as though in my own behalf, but in behalf of the other.<sup>3</sup>

Similarly, Selma Sevenhuijsen claims that care "is a cognitive and moral activity... an ability and a willingness to 'see' and 'hear' needs."<sup>4</sup>

<sup>2</sup> Metaethics concerns a cluster of questions about the nature of moral facts, moral properties, moral knowledge, moral judgment, and moral language.

<sup>3</sup> Noddings 1984, 16.

<sup>4</sup> Sevenhuijsen 1998, 82–3. How far care and compassion coincide is debatable, though. Virginia Held, for example, claims that compassion, but not care, requires altruism (Held 2006, 34–5, 42).

The rest of the introduction proceeds as follows. First, I give an outline of my argument, describing each chapter's main claims (§1). In the remaining sections, I offer some scene-setting for the argument. These sections are primarily aimed at readers who are familiar with contemporary literature on metaethics, ethics, and compassion/sympathy/empathy. Other readers should feel free to skip to the first chapter after the end of §1. For readers immersed in recent literature, however, it is important to set the scene because my approach departs from mainstream approaches in several respects. For that reason, I explain two respects in which this book might reasonably be called "old-fashioned" (§2). I then explain my choice of argumentative framework, which is broadly reductive and naturalistic, though not eliminative (§3). Following that, I locate the limits of my ambitions by describing some important debates that I remain (mostly) neutral on (§4).

## 1. Outline of the Argument

This book has three parts. Part I argues that compassion has a certain perceptual value and that, therefore, there are some aspects of reality that can be grasped only by paradigmatically good agents (at least, among agents whose desires and actions are internally coherent). Part II extends that conclusion to answer the "why be moral?" question, by considering a range of moral complexities. The arguments of Parts I and II are compatible with a range of metaethical views, including anti-realism and relativism. However, Part III uses the results of Part I to construct a form of non-relativistic moral realism by giving an account of certain moral facts, moral judgment, and moral knowledge.

The main claims of the chapters in Part I are as follows. In Chapter 1, I defend a certain approach to the "why be moral?" question, drawing on the work of Plato, William Wollaston, and Arthur Schopenhauer. The approach in question aims to show that there is a special epistemic good (i.e., a way of getting reality right) that only paradigmatically morally good people can possess. This epistemic good is special insofar as it cannot be *replaced* with other epistemic goods.

In Chapter 2, I use John Locke's philosophy of mind to illustrate the basic idea behind the answer to the "why be moral?" question I ultimately defend, which is also the core idea behind the form of moral realism I defend in Part III. Locke held that certain ideas had a special epistemic value, namely, that of resembling qualities in their objects. The paradigm for this for Locke was spatial perception, where he held that our ideas of (e.g.) shapes resembled qualities in the objects we experience. Locke did not extend this line of thought to cases of compassion, but I show how he easily could have. Moreover, Locke's theory provides resources for answering some potential objections to treating compassion as world-matching. For instance, Locke was aware that even visual perception of three-dimensional objects like spheres requires top-down processing, but shows no sign of thinking this undermines its epistemic value.

## 4 INTRODUCTION

In Chapter 3, I describe and defend the existence of a certain epistemic good. That good, which I call “being in touch,” involves representing a given object in a way that reveals some property that the object really has. To be in touch with, say, a bubble’s spherical shape would require having a representation of the bubble as somehow given (as in normal perception) in a way that reveals the property of being spherical. This is, roughly, the good of *experiencing things as they really are*. In a broadly Kantian vein, I claim that this epistemic good is not replaceable with other goods, such that (e.g.) merely reliable propositional knowledge of how things are cannot fully make up for its lack. I defend the existence and value of being in touch by appealing to both actual human echolocation and a series of contrasting fictional examples. I also distinguish it from some more familiar notions in the literature, such as “perceiving-as” and “perceiving-that.”

Chapter 4 is the core of this book. There, I aim to show (I) that compassionate people who fit one paradigm of moral goodness are in touch with the pains of others, and (II) that, among well-functioning agents (i.e., those that act in accordance with their affective states), all non-compassionate people lack that epistemic good. Among well-functioning agents, only those with compassion, I argue, experience others’ suffering in a way that reveals to them the nature of that suffering. That claim is the keystone of my answer to the “why be moral?” question: one should be moral because that is necessary for truly experiencing other creatures as they are. It also provides the keystone for the account of moral facts and moral representation I present in Part III. To illustrate why compassion is necessary for this epistemic good, I consider two other cases. The first contrasts a compassionate person with someone who has pain revealed only in his experience of himself. The second case contrasts a compassionate person with someone who is moved to avert his eyes from others’ suffering. In both cases, I argue, the compassionate person is epistemically better off than the other person. My analysis of these cases appeals to an imperatival characterization of pain and compassion (e.g., “hey you, get rid of this/that state!”).

In Part I, I focus on *short-range* compassion, that is, compassion directed at a single creature in one’s immediate environment. In Part II, I extend the argument by moving to cases involving *long-range* compassion, thereby providing a full answer to the “why be moral?” question. In Chapter 5, I argue that we can be in touch with past, current but spatially distant, and future pain states. Chapter 6 turns to pleasures and desires, and shows how long-range compassion with them is possible. Chapter 7 addresses the question of combination, such as how long-range compassion applies to scenarios involving multiple subjects with different potential desires and pains. Chapter 8 then considers the ultimate scope of long-range compassion, arguing that complete long-range compassion would be a certain God-like attitude. For each of these chapters, I argue that someone with long-range compassion fits one familiar paradigm of a morally good person.

The last two chapters of Part II address possible objections to my epistemic answer to the “why be moral?” question. Chapter 9 considers complex cases where compassion

seems to come apart from paradigmatic moral goodness, such as in feeling compassion for sadists. In each case, I argue that long-range compassion, carefully spelled out, always aligns with one familiar paradigm of moral goodness. Chapter 10 considers a more general objection, namely, why anyone should *care* about being in touch with others' pains, pleasures, and desires—glossed as a “so what?” question. My response to this is partly concessive and partly defensive. On the concessive side, I allow that some immoral agents may not care about being in touch. My answer will have no motivating force against such agents. This is a reasonable concession, however, since even a successful happiness-based answer to “why be moral?” would have no force against someone who did not care about happiness. On the defensive side, I argue that the *only* properties we can be in touch with might be others' pains, pleasures, and desires. If so, then the value of long-range compassion is increased by the scarcity of being in touch. More tentatively, I mention several controversial but familiar views according to which our minds must be directed at getting the world right, and why perception-like contact with the world must play a central role. Even if those views were right, however, someone might still object to my answer by insisting that morality is unreal or illusory in some way. Part III offers a response to this final objection.

Parts I and II are largely neutral on the nature of moral value and moral knowledge. They point the way, however, towards a ground-up account of morality. Part III uses the results of Part I to defend a novel form of moral realism, Compassionate Moral Realism (“CMR” for short). I aim to show that CMR has a number of attractive features, albeit without arguing against other metaethical views. Chapter 11 concerns the nature of moral realism generally. I consider several criteria that a view might be required to meet in order to count as moral realism, and identify three that I take to be jointly sufficient for moral realism: making some moral claims come out literally true, positing some moral facts that do not depend on anyone's “stance” towards the world, and showing that paradigmatically bad people are epistemically worse off than paradigmatically good people. These criteria help set the agenda for the rest of Part III.

Chapter 12 concerns the literal truth of moral claims and the existence of stance-independent moral facts. I propose a conceptual truth, which I call “Bad Enough,” according to which something is objectively bad if anyone who was in touch with it would feel averse to it. Bad Enough shares some of the plausibility of other analyses of value in the literature, and I argue that Bad Enough is not undermined by Moore's famous open question argument. Together with the conclusion of Part I, Bad Enough implies that it is literally true that some pains are objectively bad. This conclusion has metaphysical import, telling us how facts about objective badness could consist in facts about subjects' aversive responses. On the basis of that, I argue that CMR yields at least some stance-independent moral facts.

Chapters 13 and 14 concern moral representation and moral knowledge. In Chapter 13, I argue that CMR can explain how a wide range of moral representations maintain an “internal” connection to moral motivations. Part of CMR's explanatory potential comes from the fact that it does not draw the distinction between morally

## 6 INTRODUCTION

good people and morally bad people by appealing to their judgments. Using the analyses from Chapter 12, CMR provides an account of moral judgments according to which those judgments can be based either in emotional reactions or in purely abstract thoughts. Chapter 13 concludes with a discussion of what other writers have called “moral perception” and uses the earlier account of moral judgment to explain the “moral content” of some perceptual or experiential states.

Chapter 14 turns to the topic of moral knowledge, which some philosophers have thought is unintelligible on any form of moral realism. Using the earlier conclusions in Part III, I argue that CMR can explain some moral knowledge as a straightforward inference from a conceptual truth and a known fact about certain pain states. To show that knowledge of the latter fact is intelligible, I consider an analogous knowable fact about the contrastive similarities of color experiences. I then discuss how CMR can avoid the most obvious objections to moral knowledge concerning evolution and disagreement. At the same time, I allow that paradigmatically bad people might have something like moral vocabulary that refers to different aspects of reality than ours (such as their own desire-satisfaction). Instead of denying this, I argue, CMR can grant it, but locate the epistemic defect of paradigmatically bad people outside of judgmental knowledge. The epistemic defect is the one described in Part I: a failure to be in touch. The concluding chapter, Chapter 15, summarizes the central tenets of Compassionate Moral Realism and considers some ways in which it might be extended.

Though this argument covers a fair amount of ground, the core claim is simple: helpful, morally good, compassionate people are not dupes, since they get reality right in a way that no thoroughly selfish, apathetic, or cruel person can. With the outline of my argument in place, I now turn to scene-setting, explaining how my project differs from others in contemporary metaethics and psychology.

## 2. Old-Fashioned Metaethics

An early modern empiricist like Locke or Hume could have put together the main line of argument in this book. In some ways, it would have been easier for them to do so than for a twenty-first-century metaethicist. There are two reasons for this. First, though contemporary metaethicists still focus on the grounds of moral facts in certain senses, they have largely abandoned the sort of ambitious grounding project I undertake here. Second, the core of my argument concerns reality-matching experiences instead of true moral judgments. The latter, though, are the epistemic good that recent metaethicists have focused on.

### 2.1. *Grounding morality*

The grounding project I attempt here is of the same general type that Plato attempts in *The Republic*. The closest precedent to my particular argument is Arthur Schopenhauer’s “Prize Essay on the Basis of Morals.” I discuss the details of these comparisons in later chapters. Here, I describe the three general aims that I think define such ambitious

morality-grounding projects: (a) explaining the origin and nature of moral facts, properties, concepts, beliefs, and/or knowledge, (b) figuring out some particular moral facts, such as whether torture is wrong, and (c) answering the “why be moral?” question in a non-trivial way. Of these three aims, contemporary metaethics typically limits itself to (a). In part, this is because academic philosophy has become better at disentangling different issues and addressing one at a time, and contemporary metaethics is more or less defined in terms of (a). There are more specific reasons, though, why most metaethicists do not address (b) or (c).

A common (though not universal) contemporary view<sup>5</sup> is that facts about where morality comes from do not themselves settle ethical questions. If so, then (a) and (b) call for separate projects. Figuring out whether, say, morality comes from God, from emotions, or from rationality would be one thing, while figuring out which things are right or wrong would be something else. In principle, we could know that morality comes from God’s commands and yet know nothing about what God has commanded us to do.

Another common (though not universal) contemporary view<sup>6</sup> is that any answer to the “why be moral?” question is either trivial or absurd. A trivial answer would be one that appealed to morality itself, e.g., “you should be moral because it’s the right thing to do!” A non-trivial answer would appeal to something other than morality, e.g., “you should be moral because it will get you on television.” According to one influential argument,<sup>7</sup> it would be absurd for genuine moral philosophy to appeal to any such non-trivial answer. Presumably, neither trivial nor absurd answers would be helpful either in explaining the nature of morality or in answering moral skeptics. If those were the only answers we could give to “why be moral?” then aim (c) would be of no help to philosophers with aim (a).<sup>8</sup>

I think that these common views are over-generalizations. *Some* explanations of where morality comes from would not imply any particular moral facts. The divine-command view is a clear example. Likewise, *some* ways of answering the “why be moral?” question would be irrelevant to moral philosophy. An answer about getting on television is a fairly clear example. I believe, however, that there is an overlooked approach to grounding morality that successfully combines all three aims. The approach my argument takes—and perhaps my optimism about that approach—is therefore the first respect in which the argument of this book is old-fashioned.

## 2.2. *World-matching ideas*

The second respect in which my argument is old-fashioned concerns its characterization of our conscious experiences, their relation to the world, and their epistemic value.

<sup>5</sup> See, e.g., Dworkin 1996, McPherson 2008, Slote 2010, 107, and Enoch 2011, 41–9.

<sup>6</sup> See, e.g., Korsgaard 1996, 38–9 and Markovits 2014. <sup>7</sup> Prichard 1912.

<sup>8</sup> See Scanlon 1998, Ch. 4 for one proposal for avoiding this dilemma (as well as Scanlon 2014, 68). What I am calling “trivial” answers might still involve philosophically interesting complexities (see, e.g., Shafer-Landau 2003, 165–7).



## 8 INTRODUCTION

Simplifying immensely, here is how fashions of characterizing experiences have changed. Building on traditions going back to the ancient world, early modern philosophers like Locke held that our mental life was built out of *ideas*, understood on the model of paintings or reflections in mirrors. Like paintings or reflections, these ideas could resemble or match other things in the world. Locke and others seemed to think that world-matching ideas had some distinctive sort of epistemic value, letting us represent things “as they are in themselves.”<sup>9</sup>

Most philosophers after Locke rejected this theory of world-matching ideas. In the twentieth century, the “linguistic turn” in philosophy involved characterizing the mind by appealing to non-imagistic, linguistic items like sentences. This latter approach survives today in the extensive use of *propositional attitudes* to frame philosophical issues. Propositional attitudes include the belief that it is raining, the desire that my finger not be scratched, the knowledge that torture is wrong, and (more controversially) the perception that someone is sad. Though propositional attitudes may not themselves require language, the principal way of characterizing them uses declarative sentences like “it is raining,” and propositional knowledge can be roughly understood as whatever is required to reliably pass written exams. Many twentieth-century philosophers understood their task as analyzing the constituents of propositional attitudes (i.e., concepts) or as giving a semantics for certain important terms. Much of twentieth- and twenty-first-century metaethics has therefore focused on moral judgments and moral concepts.<sup>10</sup> Even metaethicists who think that empathy, compassion, or other emotional responses are central to morality tend to focus on how those capacities *contribute* to our making moral judgments.<sup>11</sup> These discussions seem to assume that the main epistemic value relevant to moral philosophy is the value of propositional knowledge. By “epistemic value,” I mean the type of value involved in getting reality right and correctly representing how things are, as well as types of justification or virtue connected to getting reality right (some contemporary philosophers use the term “epistemic” more narrowly than this).

My own background is largely in seventeenth- and eighteenth-century philosophy, and I have always felt more comfortable characterizing the mental in terms of ideas than in terms of propositional attitudes. Through the efforts of some patient teachers, though, I came to understand that we cannot characterize our mental life *entirely* in terms of image-like ideas, and that many aspects of the mind are better understood in terms of propositional attitudes. A key premise of the argument I make below, however, is that one important and epistemically valuable dimension of our experience

<sup>9</sup> See *Essay* 2.28.23, discussed in Chapter 2.

<sup>10</sup> Kwame Anthony Appiah gives a similar diagnosis: “I suspect that a preoccupation [in moral philosophy] with whether moral sentences are ‘fact-stating’ or ‘truth-apt’ is, in some measure, an artifact of the exaggerated authority that the philosophy of language . . . has enjoyed for much of the postwar era” (Appiah 2008, 181–2).

<sup>11</sup> For instance, Prinz 2007, Slote 2010. For one significant exception, albeit one only partly concerned with compassion, see Arpaly 2002 (especially 92–3). Arpaly’s view of the limitations of moral judgment aligns with some of the claims I make in Chapter 13.



is best captured on the model of reality-matching ideas, and that there is no obvious way of capturing this in terms of propositional attitudes (though there may be some non-obvious way of doing so). This is a broadly Kantian idea—emphasizing the distinctive importance of what is “given” in experience, in contrast to our judgments and statements about what is given. Ernst Cassirer offers a dramatic expression of this idea:

If language is to grow into a vehicle of thought, an expression of concepts and judgments, this evolution can be achieved only at the price of forgoing the wealth and fullness of immediate experience. In the end, what is left of the concrete sense and feeling content it once possessed is little more than a bare skeleton.<sup>12</sup>

Because of this, the first two parts of this book say little about moral concepts and the meaning of moral terms. I turn to these issues in Part III, but they are neither my starting-point nor my main concern. Unlike many other recent philosophers who have discussed compassion and metaethics, I do not claim that compassion is necessary for moral judgments, or even for moral knowledge. Cold sociopaths, I think, could have any amount of test-passing propositional moral knowledge. Even so, however, they would be epistemically worse off than compassionate people in another respect.

This focus on world-matching ideas and their value is the second respect in which the argument that follows is old-fashioned. It is also, I think, the main reason why the general approach I take to grounding morality here has not been explored by recent philosophers.<sup>13</sup>

### 3. Framework: Non-Eliminative Naturalism

I now turn to explaining some of my guiding assumptions about what a proper grounding of morality would look like. What I say here primarily concerns the view I defend in Part III, but it also bears on the metaphysical assumptions I make in Parts I and II. Among the many broad brush-strokes of this introduction, this section contains some of the broadest.

Some facts might have no deeper explanation. Logical laws (e.g., if A, then A) seem like good candidates for such facts. It seems worth *trying* to give some deeper explanation for logical laws, but it also would not be surprising if they were the endpoints of explanations, i.e., brute facts. This is connected to our sense that logical laws hold *necessarily*. That is, it does not seem that the laws of logic could have been different or could change. Rules of grammar (e.g., every declarative sentence should contain a verb) seem to differ from logical laws in both these respects. Rules of grammar seem to obviously have some deeper explanation. This is obvious even to those of us who know

<sup>12</sup> Cassirer, 1925/1946, 98.

<sup>13</sup> The closest contemporary precedents to my argument, albeit on different points, are the views proposed by Christoph Fehige (Fehige 2001, Fehige 2004), Justin Steinberg (Steinberg 2014), and Kate Manne (Manne 2016, Manne 2017)—though none of these authors, I think, sees the potential of understanding compassion as perception. I note some relevant points of comparison below.

## 10 INTRODUCTION

almost nothing about linguistics, and have no real idea of how the details of that explanation might go. Connected to that, the rules of grammar do not seem necessary in any important sense, and so could easily change.

Autobiographically, it has seemed to me for a long time that morality is an intermediate case between logical laws and rules of grammar. The fact that there is something bad about torture, for instance, seems like a necessary truth, and so more like a logical law than a rule of grammar. I cannot imagine a scenario in which there would be *nothing* bad about torture. On the other hand, the badness of torture does not seem like a brute fact. Like a rule of grammar, it seems like it has some explanation, even if we are unsure about the details of that explanation.

The analogy with rules of grammar may go further. We non-linguists do not know what the explanation of grammatical rules is, but we have a general sense of *where* that explanation might come from: facts about the brain, social development, and the environments of our evolutionary ancestors. We would be surprised to hear that grammar can be explained only by appeal to supernatural properties. Put another way: the existence of evolved social creatures with brains like ours should be enough to explain rules of grammar. Since I started worrying about metaethics, something similar has seemed true to me concerning morality. It seems that an explanation of morality should come from facts about conscious suffering, happiness, feelings of compassion, and certain familiar features of rational agency. It would be surprising if morality could be explained only by appeal to supernatural properties. Put another way: the existence of creatures who can suffer, be happy, feel compassion, and reason should be enough to explain the facts of morality.

My metaethical inclinations differ from those of most early modern philosophers. René Descartes, for instance, held that grammatical abilities could be explained only by appeal to a non-physical soul, and that moral facts (and perhaps logical laws as well) were explainable only by appeal to a supernatural God.<sup>14</sup> David Hume, by contrast, held that moral rules could be explained in terms of human nature, but denied they were necessary in any deep way.<sup>15</sup> Except perhaps for marginal figures like William Wollaston (discussed in Chapter 1), no early modern ethicist that I know of clearly endorsed the necessary-but-non-supernatural view of morality.<sup>16</sup>

Many contemporary philosophers hold that morality is not necessary in any significant way, while others hold that it is either brute or explainable only by appeal to something non-natural.<sup>17</sup> Some of these contemporary philosophers might just find it

<sup>14</sup> See, e.g., *Discourse*, Part 5 on language and non-human animals and Meditations 3 and 4 on our ideas of God and of imperfection.

<sup>15</sup> See, e.g., *Treatise* 3.1.1–2.

<sup>16</sup> I say “clearly” because there are ways of reading some early modern philosophers (like Spinoza and Kant) that would bring them closer to the necessary-but-natural view.

<sup>17</sup> An influential example of the former is Harman 1977, while an influential example of the latter is Nagel 1986. For (apparent) rejections of necessary-but-natural intuitions like mine, see Huemer 2005, 94 and Enoch 2011, 106–8. Some self-described “non-naturalist” moral realists have related intuitions, however (e.g., Shafer-Landau 2003, 74–7). As Shafer-Landau notes, taking moral facts to consist in familiar facts

obvious that moral facts are not necessary, or that they are brute, or that they involve something non-natural. Others, however, might share my initial inclinations, and yet think that larger considerations force us to give up something. Among contemporary philosophers, my main audience is this final group. That is, I hope to show how morality might be necessary and yet explainable in terms of familiar mental features.

In contemporary jargon, my overall approach is broadly naturalistic and reductive, but non-eliminative. It is broadly naturalistic in that it does not assume the existence of anything other than entities recognized by contemporary cognitive psychology. I say “broadly,” since I remain neutral on whether conscious states and a modest conception of rationality can be explained in purely physical terms. In addition, my claim that reality-matching experiences are a sort of epistemic good is not entailed by anything empirical, and makes an epistemic normative claim.<sup>18</sup> The view I ultimately defend in Part III is reductive, but non-eliminative, in that it aims to provide an (analytic) explanation of morality in terms of non-moral things, but without suggesting we should eliminate morality from our objective picture of the world.<sup>19</sup> By analogy: it seems we could exhaustively explain (and so “reduce”) facts about Oregon’s economy in terms of facts about humans and their social relations without thereby suggesting that we should eliminate Oregon’s economy from our objective picture of the world.

## 4. Fights I Am Not Picking

My project is a positive one: offering a broadly naturalistic grounding for morality, by way of an appeal to how compassionate reactions can match reality. Pursuing this project requires taking sides in a number of fights, especially with certain sorts of scientifically minded skeptics about morality. I am not particularly interested in competing with other positive accounts of morality and compassion, however. For that reason, I want to mention a number of well-known debates that I am not addressing here, at least not in any direct way. These other debates all concern important issues, and I do hope that my argument can shed some light on them, but you should not expect any in-depth discussion of them below.

### 4.1. *The nature of empathy and compassion*

The first group of debates I am not taking sides on concerns compassion. Some of these debates are terminological, while others are more psychological and philosophical.

about people’s actions and mental states can explain why the moral supervenes on the natural, that is, why situations cannot differ morally without also differing in some natural respect.

<sup>18</sup> This normative element does not significantly undermine the view’s naturalist credentials, however, unless we thought that epistemic normativity cannot be understood naturalistically. For reasons given in Chapter 10, the epistemic normativity I appeal to is quite lightweight, and I see no reason to think it is non-natural.

<sup>19</sup> The view I defend in Part III is analytic in that it aims to link moral facts to non-moral facts through an analysis of moral concepts. It differs significantly, however, from other recent forms of analytic naturalism (such as Jackson 1998).

## 12 INTRODUCTION

My argument focuses on how some conscious beings' feelings can match other conscious beings' feelings. The term I use for this phenomenon is "compassion," even though the term "empathy" might also fit ("sympathy" is also a reasonable contender). I make this terminological choice because there is a large empirical and philosophical literature discussing empathy, most of which focuses on slightly different phenomena than the one I am concerned with. The term "compassion" happens to be less common in the recent literature (though, by the measure of a recent Google search, it is more than twice as common overall), so using "compassion" helps avoid some misleading appearances of disagreement.

The phenomena that other writers have applied these terms to all involve psychological states that are somehow linked to other subjects' psychological states. Often, but not always, these are states with some sort of emotional or affective dimension. The terms "empathy," "sympathy," "compassion," and "fellow-feeling" have been applied to everything from simple emotional contagion (e.g., an infant becoming distressed because another infant is crying) to deliberate inner simulations of other minds (e.g., imagining how Socrates felt when he was sentenced to death) to certain shared emotional states (e.g., grieving together at a loss)<sup>20</sup>. Which phenomenon is discussed often depends on what broader issues are under discussion, such as moral and social development, moral judgment, moral virtue, or our (not necessarily moral) knowledge of other minds.<sup>21</sup>

One reason why my focus differs from that of many other recent writers is the following. Many philosophers and psychologists characterize compassion, empathy, or sympathy "from the inside." On this approach, intrinsic features of a creature are sufficient to determine whether a creature is demonstrating compassion (empathy, etc.). For example, if it seems to you that someone is distressed, and that makes you feel distressed for them, many people would say that you are demonstrating compassion (empathy, etc.), even if that other person is not actually distressed, or even if you are hallucinating that other person.

By contrast, the phenomenon I call "compassion" is characterized both "from the outside" as well as from the inside. This characterization comes from my focus on states that match or resemble other states. If you feel distressed in response to a mere hallucination, then your distress does not match the state of any real creature outside you. In that case, you would not be demonstrating compassion in my sense, even if you might justifiably believe that you were—and, in fact, even if only God could conclusively determine that there was no match. The inside/outside distinction here is parallel to an ambiguity in ordinary talk of perception. In one sense of "perceives," if it seems to you that you are seeing a cat, then you are perceiving a cat. In another sense, though, it is not enough for it to merely seem that way to you—there must also be a cat there. My

<sup>20</sup> Max Scheler counts this last case as the "highest form" of fellow-feeling (*Mitgefühl*) (Scheler 2009, 12–13).

<sup>21</sup> For helpful overviews, see the introduction to Coplan and Goldie 2011 and Chapter 1 of Batson 2011. My notion of compassion comes close to the definition of empathy in de Vignemont and Jacob 2012.

use of “compassion” is like this second, success-requiring sense of “perceives,” whereas much of the literature uses the term in the first, success-neutral sense.<sup>22</sup>

There is another respect in which the phenomenon I focus on differs from related phenomena. Many philosophers and psychologists take the primary case of compassion, empathy, or sympathy to be that of engaging with a *single present* mental state of some creature.<sup>23</sup> In such a case, the match involved in compassion would be between the present state of the compassionate person and the present state of the creature that person has compassion for. While I consider this kind of case in the early chapters of this book, my ultimate interest is in a mental capacity that can match multiple states, where those states can be in the past, in the future, or be merely possible. I describe this contrast using a distinction between “short-range” and “long-range” compassion. Some familiar worries about connecting morality to compassion, it turns out, concern only short-range compassion.<sup>24</sup> It is long-range compassion, however, that I think provides a grounding for morality, though the relevant facts about matching are (not surprisingly) more complicated than with short-range compassion.<sup>25</sup> One important complication is that a long-range compassionate (and so morally good) person need not have the same type of emotions as the present objects of her compassion. For example, if there is no real danger, a long-range compassionate response to fear need not involve fear.

Finally, some philosophers and psychologists focus on phenomena that involve “conceiving what we ourselves should feel” in someone else’s situation, or “changing places in fancy with the sufferer.”<sup>26</sup> While compassion, in my sense, often draws on background knowledge of one’s own experience, it does not require any thought of oneself. One may be pained by another’s suffering immediately, without considering whether one would feel the same in the other’s situation—or even while being firmly convinced that one would *not* feel the same.

<sup>22</sup> For example, Stephen Darwall writes, “Empathy consists in feeling what one imagines [another] feels, or perhaps should feel (fear, say), or in some imagined copy of these feelings” (Darwall 1998, 261). Michael Slote argues that we can have “empathetic concern for a group that one merely hears about that may not, in fact, exist” (Slote 2010, 39). By contrast, some psychologists employ a notion of empathetic accuracy, though this only partly aligns with the notion I employ (see, e.g., Ickes 1997). In recent philosophy, an appeal to outside success conditions for empathy is found in Lori Gruen’s notion of “entangled empathy,” which has both epistemic and ethical success conditions (Gruen 2015, 83–94; see also Piper 1991, 738, Deonna 2007, and Fehige 2004 on “a priori empathy,” but cf. Nussbaum 2001, 328).

<sup>23</sup> Martin Hoffman, for example, claims that empathy has a “here-and-now bias” (see Hoffman 2000, Ch. 8), though he claims that it typically develops beyond this.

<sup>24</sup> For a recent example, see Bloom 2016, 98.

<sup>25</sup> Most philosophers who see a connection between compassion and morality appeal to something like long-range compassion. See, e.g., Hume, *Treatise* 3.3.5.1, Smith, *Theory* 1.1.4.8, Sidgwick 1907/1981, 455, 500, Slote 2010, 106, 138, Bein 2013, 91. For a relevant general discussion, see Kauppinen 2014. In addition, several religion traditions connect God’s goodness to God’s infinite compassion (see Chapter 8).

<sup>26</sup> Adam Smith is one example (the quoted phrases are from *Theory* 1.1.1.2–3). A related idea, as with Aristotle on pity (*Rhetoric* 1385<sup>b</sup>–1386<sup>b</sup>), is that compassion/sympathy/empathy requires thinking that something similar could befall oneself.

## 14 INTRODUCTION

In using the term “compassion” for one phenomenon among others, I am not claiming that my use of “compassion” is more familiar, natural, or widespread than other uses. I do suspect, though, that my use is at least somewhat familiar. Most people, I suspect, would be reluctant to call someone “compassionate” whose states *never* matched others’ states. Likewise, I suspect most people would be willing to talk of feeling compassion for others’ future suffering. Regardless, how these terms are used in everyday English does not ultimately matter for my argument. I need a term, and “compassion” seems as good a choice as any. This choice should not be confused with a substantive claim about the nature of some phenomenon we have previously identified or about ordinary language use.

Terminological issues aside, I also want to avoid a number of psychological questions about compassion. This is not because I want to avoid psychological issues entirely; Chapters 3 and 4 appeal to the psychological literature on several points. What matters most for my purposes is that we can have certain motivational states that match other motivational states, and that this ability is similar to certain perceptual processes. For the most part, it does not matter for my argument how compassion develops in humans or under what circumstances we succeed or fail in showing compassion. Nor does it matter how common long-range compassion is, so long as it is possible for us to have it to some degree. Though these things do not matter *for my argument*, my argument implies that they matter tremendously. In particular, the way that various biases interact with our compassionate tendencies is deeply troubling, and we could become morally better as a species if we could develop truly unbiased compassion. My hope is that the argument I give here helps us say what truly unbiased compassion would be and why it matters. Saying that, however, requires relatively little discussion of our psychology, just as saying what a truly corruption-free government would be and why it matters does not require saying much about the actual causes of corruption.

Finally, there are two related philosophical debates that I largely avoid. The first is about whether emotional reactions like compassion presuppose some belief about value, or have moral “content.” For example, Martha Nussbaum, following Aristotle, claims that compassion requires believing that the suffering one is focusing on is both serious and undeserved.<sup>27</sup> In what follows, I write as though this were not the case, that is, as though compassion can occur in someone who has no conception of value at all. My reason for doing so is that, in Part III, I aim to build an account of moral value and moral beliefs on the basis of non-moral facts. Nonetheless, the argument in Parts I and II is compatible with saying that compassion somehow contains an evaluation, so long as there is a relevantly similar (matching) evaluation in the suffering the compassion is directed towards. That kind of possibility would complicate my overall story in some

<sup>27</sup> Nussbaum 2001, 306; Aristotle *Rhetoric* 1385<sup>b</sup>–1386<sup>b</sup>. In a related vein, Daniel Elstein and Thomas Hurka suggest that the concept of compassion involves a moral component (“sorrow at another’s evil”—Elstein and Hurka 2009, 529). Note that Nussbaum uses “empathy” to describe something closer to what I am using “compassion” for. She denies that empathy requires evaluative beliefs (Nussbaum 2001, 328–9). Max Scheler takes a similar line (see Scheler 2008/1913, 5). See also Nagel 1970, 80 on sympathy.

respects, but not undermine it.<sup>28</sup> For that reason, I am not arguing against those who hold that compassion presupposes evaluative beliefs or has moral content, though I discuss some related issues in connection with moral judgment in Chapter 13.

The second philosophical debate I want to avoid is about the ultimate nature and content of compassion and related states like pain. At several points, I appeal to facts about the brain to provide support for claims about these states. However, the argument I make is consistent with those states having all sorts of metaphysical statuses. They might be irreducibly mental, functional, or reducible to physical states. In Chapter 4, I offer a characterization of these states in terms of imperatives (e.g., “change this!”), but I do not claim that this characterization captures the ultimate nature of those states. My argument is consistent with these states having conceptual, non-conceptual, propositional, or non-propositional content, or having no content at all. What matters for my purposes is that these states are distinct from other states, in particular from familiar types of propositional knowledge.

#### 4.2. *Other approaches to grounding morality*

I have described my aim as grounding morality, which suggests that morality has not already been grounded. I would not have written this book if I had thought we already had a fully satisfactory account of the grounds of morality, but it is not part of my aim to argue against other positive metaethical views. For one, my answer to the “why be moral?” question is consistent with a wide range of views about the nature of moral facts and moral knowledge. For another, the metaethical view I offer in Part III might be only part of a full metaethical picture.

For example, some people think that morality is grounded in facts about God or about our rational capacities. Others think that morality is a robust, independent feature of reality, and so needs no additional grounding. I do not think that we have a fully satisfactory version of these approaches yet. My doubts about them (hinted at in §3 above) are already well expressed by others, though, and are separable from my positive argument. For that reason, I do not attack any other optimistic views about morality. This raises a question, though: could *more than one* approach to grounding morality be correct? For anyone who thinks morality is already well-grounded in some other way, I offer my main argument here as a supplement or reinforcement. If, contrary to what I have assumed, moral facts are real but need no grounding, then perhaps what I describe below is only indirectly related to morality (if there is a better account than mine, it can have exclusive rights to the term “morality”). On the other hand, it may be that the different grounds could converge in some way. In that case, what I offer here might be only part of the picture, and morality might turn out to involve different elements.<sup>29</sup>

<sup>28</sup> For a view that appeals to resemblance and evaluative states in something like this way, see Montague 2014. See Appendix C for more discussion.

<sup>29</sup> As Mark Schroeder writes, “when some philosopher tries to give us an account of why we ought to be moral, he usually gets at least *something* right” (Schroeder 2007, 116).



For these reasons, I am only interested in a limited fight about the nature of moral concepts and moral terms. In Part III, I offer an account of some moral concepts, but my aim there is only to show that the argument of Part I can be developed into one form of what is today called “moral realism.” Nothing in Part III is intended to rule out other forms of moral realism or other (similarly optimistic) accounts of moral concepts or terms.

#### 4.3. *Controversial moral cases*

The third and final class of debates I stay neutral on concerns certain controversial moral facts. The first part of my argument does take a stand on a number of specific moral comparisons. For example, I claim that someone who is pained by the thought of another creature’s future pain is, other things being equal, morally better than someone who is pleased by that same thought. This, however, is a relatively uncontroversial case. For my argument to succeed, it needs to line up with *many* of our views about uncontroversial moral cases. More specifically, I argue below that compassionate people (in my sense of “compassionate”) are people who we would count as being morally good. The simplest paradigm of such a person, in my discussion, is someone who is pained by the sight of another creature’s suffering and thereby moved to help (I discuss the connection between being pained and being moved in Appendix A). Few ethicists would deny that pain and suffering are bad, or that (other things being equal) good people help alleviate others’ pain.

Even so, there are other widely recognized paradigms of moral goodness or virtue. One example is the righteous warrior, who rages against injustice and punishes the wicked. Another example is the absolute loyalist, who never lies, never breaks a promise, and never lets down a friend or family member. A third example is someone who perfectly cultivates her own character. None of these paradigms requires compassion in any direct way, so the view I defend below does not ground their (at least apparent) moral goodness or virtue.<sup>30</sup> If the view I lay out by the end of Part III gives the complete story of morality, then we should question whether these other paradigms are well-founded.

To me, these other moral paradigms do seem questionable. I suspect that some views about retributive justice come from mistaken beliefs about free will, that some of the appeal of absolutist principles comes from a sort of deliberative laziness, and that focusing on self-development is unacceptably self-indulgent. I do not argue for those views here, though. As I mentioned above, for all I say here, it could be that morality has more than one ground. If so, different considerations from those I provide might vindicate the righteous warrior, the absolute loyalist, and the self-cultivator. When I write of giving a ground of morality below, then, this should always be understood as leaving room for other grounds.

<sup>30</sup> For a discussion of one intellectual tradition opposed to compassion, see Nussbaum 2001, Chapter 7.

All of this means that my argument stays neutral on a wide range of controversial moral cases, especially cases that involve a conflict between these different paradigms. Sometimes relieving suffering precludes punishing the wicked, keeping a solemn promise, or pursuing one's own development, for example. I argue below that, in such cases, there is something important to be said on behalf of relieving suffering, but that leaves open that there is also something to be said on behalf of retributive punishment, absolute loyalty, or self-cultivation.

Finally, I should emphasize that nothing in my account implies that all people who lack compassion are morally *bad*. It in fact follows from my larger view (as developed in Part III) that people lacking in compassion can both have full-fledged moral knowledge and, moreover, be uniquely suited to make the world morally better in some ways. Some morally horrific situations can cause trauma and debilitating compassion-fatigue in compassionate humans, after all—our emotional capacities have narrow limits. In such situations, the people who can do the most good might be those who do not emotionally engage with others.

With all that in place, I now turn to my core argument. My first aim is to show that compassionate people, like those who helped me escape the subway doors, get reality right, while all sadistic and indifferent people are missing something.