

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

CHARACTER AND MORAL JUDGMENT: DESIGNING RIGHT AND WRONG

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

SETH ROBERTSON

Norman, Oklahoma

2019

CHARACTER AND MORAL JUDGMENT: DESIGNING RIGHT AND WRONG

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF PHILOSOPHY

Dr. Nancy Snow, Chair

Dr. Martin Montminy

Dr. Amy Olberding

Dr. Linda Zagzebski

Dr. Carolin Showers

© Copyright by SETH ROBERTSON 2019
All Rights Reserved.

For my parents

Acknowledgements

It is impossible to acknowledge here each person to the extent they deserve in (directly or indirectly) helping me write this dissertation. First, I must thank my professors at Juniata College, especially Xinli Wang, Jack Barlow, and Brad Andrew for providing a truly transformative intellectual and moral experience. I also must thank my professors and fellow graduate students at the University of Houston for teaching me how to begin to do academic philosophy. And I must thank a host of people at the University of Oklahoma and in the Philosophy Department here for the tremendous amount of great help, advice, support, mentoring, feedback, and instruction that they gave. In particular, I must thank all the members of my dissertation committee, who provided very much and very useful feedback on this dissertation. Nancy Snow, my advisor, gave an incredible amount of feedback (and encouragement) over an exceedingly high number of drafts of an exceedingly large range of polish and quality. Chapter 3 of this dissertation benefited from feedback from Matt Priselac and Max Parish. The project in general benefitted greatly from conversations with my colleagues at the Institute of Human Flourishing last year: Jing Hu, Alex Danvers, Jessica Black, and Jong-hyun Jung. I also must thank the Department of Philosophy for creating and maintaining an environment that supports graduate student success: in particular must thank Steve Ellis and Wayne Riggs (and Hugh Benson) as well as Shelley Konieczny and Gabe Serrano on this front. I also would like to thank Amy Olberding in particular for providing a huge amount of excellent feedback and advice during my time at the University of Oklahoma. This dissertation was made possible through the support of a grant from the John Templeton Foundation to the Institute for the Study of Human Flourishing at The

University of Oklahoma. The opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation, the Institute for the Study of Human Flourishing, or The University of Oklahoma.

Table of Contents

| | |
|---|----|
| Acknowledgements | iv |
| List of Figures..... | ix |
| Abstract..... | x |
| Chapter 1: Introduction..... | 1 |
| 1.1 Introduction | 1 |
| 1.2 Implications of the Consequences and Character Conditions | 6 |
| 1.3 The Argument..... | 9 |
| 1.4 The Plan..... | 14 |
| Chapter 2 | 14 |
| Chapter 3 | 14 |
| Chapter 4 | 15 |
| Chapter 5 | 16 |
| Chapter 2: Overview..... | 17 |
| 2.1 Introduction | 17 |
| 2.2 The Consequences and Character Conditions | 17 |
| 2.2.1 The Consequences Condition | 19 |
| 2.2.2 The Character Condition | 23 |
| 2.3 Perspectival Naturalism..... | 31 |
| 2.3.1 Naturalism / Descriptive adequacy..... | 35 |
| 2.3.2 Anti-realism..... | 37 |
| 2.3.3 Non-Nihilism / Constructivism | 38 |
| 2.3.4 Perspectivalism..... | 39 |

| | |
|---|-----|
| 2.3.5 The Design Paradigm | 40 |
| 2.4 Summary..... | 43 |
| Chapter 3: The Case for Perspectival Naturalism | 44 |
| 3.1 Introduction | 44 |
| 3.2 Radical Evolutionary Debunking | 47 |
| 3.2.1 Traditional Debunking Arguments..... | 47 |
| 3.2.2 The Radical Debunking Argument..... | 53 |
| 3.2.3 Responses to Debunking Arguments..... | 64 |
| 3.3 Constructing Morality | 71 |
| 3.3.1 Metaethical Constructivism..... | 71 |
| 3.3.1 Perspectival Naturalism..... | 75 |
| 3.3.2 The Design Paradigm | 80 |
| 3.3.3 Perspective..... | 98 |
| 3.4 Summary..... | 105 |
| Chapter 4: The Consequences and Character Conditions | 107 |
| 4.1 Introduction | 107 |
| 4.2 The Consequences Condition | 107 |
| 4.2.1 The Never-Consequences Condition..... | 108 |
| 4.2.2 The Always-Consequences Condition | 115 |
| 4.2.3 Summing Up: The Argument for the Consequences Condition..... | 117 |
| 4.3 The Character Condition | 118 |
| 4.2.1 Person-centered accounts of moral judgment and the practical perspective | 119 |

| | |
|---|-----|
| 4.2.2 Goodness in actual and possible consequences | 124 |
| 4.2.3 Individual Differences in Character | 130 |
| 4.2.4 Value and Absurdity | 139 |
| 4.3 Objections to the Character Condition | 144 |
| 4.3.1 The Person-Neutral Objection | 144 |
| 4.3.2 The Situationist Objection and the Responsibility Objection | 147 |
| 4.3.3 Reasons and Excuses | 155 |
| 4.4 Summary | 156 |
| Chapter 5: Conclusion | 158 |
| 5.1 In Summary | 158 |
| 5.2 Implications reconsidered | 164 |
| 5.3 Directions for future research | 177 |
| References | 186 |
| [Appendix A: Truth, Relativism, and Pluralism] | 196 |

List of Figures

Figure 1: The Frankfurt Kitchen (view from the entrance). Public Domain.

https://en.wikipedia.org/wiki/Frankfurt_kitchen#/media/File:Frankfurterkueche.jpg ... 86

Abstract

I argue that an adequate theory of rightness should meet (at least) two distinct conditions: a “Consequences Condition” according to which the rightness or wrongness of some, but not all acts should be determined *conclusively* by the act’s outcomes on welfare, and a “Character Condition” according to which the rightness or wrongness of some, but not all acts should be influenced *partially* by aspects of the moral character of the person who committed the act. The combination of these two conditions is interesting because many major normative theories capture one, but not both well. In the course of making the case for the Consequences and Character Conditions, I develop and argue for a novel version of metaethical Humean Constructivism that I call “perspectival naturalism,” which I then apply in support of the Consequences and Character Conditions.

Chapter 1: Introduction

1.1 Introduction

Puzzles about rightness and wrongness have long bedeviled moral philosophers.

Consider the following cases, based on an example from Sverklík (2011):

Mary’s Anxiety: Mary is at a dinner gathering with friends, and is introduced to several new people, including Arjun. Mary has some social anxiety and feels nervous shaking hands with people, so instead of shaking hands with Arjun, she waves. No one else (including Arjun) notices or is bothered.

Mary’s Prejudice: Mary is at a dinner gathering with friends, and is introduced to several new people, including Arjun. Mary has an extreme prejudice against people from India, and so she refuses to shake Arjun’s hand. Instead, she waves. No one else (including Arjun) notices or is bothered.

In both cases, we might imagine that Mary’s bodily movements, facial expressions, gestures, and tone of voice are identical, and since in neither case anyone else really noticed or was bothered by Mary’s actions, the immediate consequences were also identical. While many moral philosophers have claimed that the motivations driving an act should have no bearing on its rightness or wrongness (e.g. Mill 1979, 17–18), many people will have the intuition that Mary did something *wrong* in the second case but not in the first *because of* her motivations.¹

¹ In the opening pages of *The Right and the Good*, W.D. Ross gives what seems to me to be a very poor argument (in an otherwise very good book) that motivations cannot justifiably be included in a theory of rightness. The argument is that since an act is right just if it is morally obligated, and we cannot be obligated to feel motivations (because this would force us to feel motivations we do not, and thus violate the “ought implies can” principle), then motivations cannot help determine what is right. Whatever one thinks about whether motivations should play a role in a theory of rightness, this argument has problems: even if we grant (and we shouldn’t) that this is an acceptable use of “ought implies can” and that we cannot become motivated when we aren’t, this argument would show that nearly nothing at all could be right, since if I cannot be motivated to do something, then I cannot do it, and if I cannot do it, then it is not the case that I ought to do it.

And of course, the question of motivation aside, there are even deeper questions and disagreements about rightness and wrongness. For example, there seem to be at least three ways that rightness is treated most generally:

Desirability: An act is right if it is the most morally desirable act for a (any?) person to perform.²

Obligation: An act is right if it is the act that a (any?) person is morally obligated to perform.³

Reasons: An act is right if it is the act that is, on the balance of reasons, the act with the best reasons to perform.⁴

This additional disagreement aside, there is still further disagreement about the relative importance of rightness for morality. A great many philosophers have been implicitly or explicitly less focused on rightness and wrongness as central moral concepts than other moral concepts: Michael Slote includes Plato, Aristotle, Abelard, Schopenhauer, Kant, Hutchinson, Nietzsche, Spinoza, and “certain Chinese and Indian thinkers” in this group (Slote 2001, 7–10). Even more strongly, Elizabeth Anscombe famously argued that we should (*at least temporarily*) discard the notions of rightness and wrongness altogether (Anscombe 1958).

There are even further a great number of smaller and more specific disagreements about rightness. For example, philosophers have argued about whether rightness should be construed as a scalar or binary notion: that is, whether rightness or wrongness come in degrees, or whether all acts are either simply right or wrong (e.g. Peterson 2013; Brown 2016). To take a final example, another important disagreement centers around whether rightness should be agent-neutral or not: that is, whether what is

² This is the sense of rightness which, for example Arpaly (2003) emphasizes.

³ This is the sense of rightness emphasized by, for example Ross in *The Right and the Good* (1930).

⁴ This is the sense of rightness utilized, for example, in Pettit’s *The Robust Demands of the Good* (Philip Pettit 2015).

right to do in a given situation would be right for *anyone* in that situation. Nel Noddings, in *Caring* (2003), explicitly denies this: some of our relationships reshape what we should do in certain circumstances.

In this dissertation, I am particularly interested in questions about the relation between our judgments of right and wrong and the moral character of the person who performed those acts. Let's begin by examining pairs of cases and asking whether the actions performed in each pair were equally right or wrong, and if not, what made it so.

Bully Bryan: Bryan is a bully. He takes genuine joy in insulting his fellow high school students in the most hurtful ways he can imagine. One day, feeling anger, hatred, disgust, and maliciousness, he insults his classmate Ivan in the cruelest way he can imagine. Ivan did not previously know Bryan, and (for other reasons) transferred schools the following day and never saw him again. Bryan continues to be a bully.

Regular Ryan: Ryan is not a bully – in general he is a typical and friendly high school student. One day, he was in a bad mood, and though it was out of character, he insulted his classmate Ivan. In that moment, Ryan felt the exact same emotions to the exact same degrees as Bryan had, and uttered the exact same insult as Bryan did. Ivan did not previously know Ryan, and (for other reasons) transferred schools the following day and never saw him again. Ryan did not bully other people after this incident.

Kind Chris: While waiting in line at the cash register in a grocery store, the person currently paying is making a small purchase (a few pieces of fruit) and their credit card is unexpectedly declined. The cashier apologizes, saying the system just crashed. The customer checks their wallet and is mortified to realize they don't have any cash, and thus cannot pay for their purchase. Chris notices all this, and without thinking about it, immediately offers to pay for the other customer.

Calculating Caleb: While waiting in line at the cash register in a grocery store, the person currently paying is making a small purchase (a few pieces of fruit) and their credit card is unexpectedly declined. The cashier apologizes, saying the system just crashed. The customer checks their wallet and is mortified to realize they don't have any cash, and thus cannot pay for their purchase. Chris notices all this, and while it is unfolding, deliberates about what he should do. He eventually comes to the conclusion that he should offer to pay, and he does, but only after weighing several moral and prudential considerations (such as that helping would make him feel good later, that it would make others think more highly of him, that he could brag about it). Ultimately, his main motivation was

to help the distressed customer, but his action was certainly not immediate or motivated *solely* by helpfulness.

Honest Amanda: Amanda is a high school student, and yesterday was the annual skip day. She skipped class, and had a fun day with her closest friends. Later that week, Amanda's mother pointedly asks whether Amanda skipped school. Amanda knows that she might get in trouble, but she is an honest person. Though it is not an easy situation, and if she lied, she likely could get away with it, she tells the truth.

Truth-telling Miranda: Miranda is a high school student, and yesterday was the annual skip day. She skipped class, and had a fun day with her closest friends. Later that week, Miranda's mother pointedly asks whether Miranda skipped school. If Miranda thought she *could* get away with it, she would not tell the truth. However, in this circumstance, she suspects that she would not be able to get away with the lie, and so she tells the truth.

Caring Kate: Kate has been accepted both to medical school and veterinary school. She believes (correctly) that if she went to medical school, she would be able to help a great many more people than the number of animals should be able to help as a veterinarian. But, she is deeply committed to animal welfare. She decides to become a veterinarian. She has a long career as a veterinarian and helps a great many animals.

Capricious Cat: Cat has been accepted to both medical school and veterinary school. She believes (correctly) that if she went to medical school, she would be able to help a great many more people than the number of animals should be able to help as a veterinarian. But, on a whim, she decides to become a veterinarian. She has a long career as a veterinarian and helps a great many animals – in fact, she helps an identical number of animals to an identical degree as Kate.

Now, I expect people will have quite different responses to the question of whether each pair of cases contains acts that are equally right or wrong. I suspect some will say that the acts are obviously equally right or wrong, while others will say that the pairs of acts obviously differ in their rightness and wrongness. Some might offer further analysis. Of course, some might try to find ways in which the consequences of the acts will diverge – and to these readers I will stipulate that the details of all cases should be filled out so that the consequences of the paired acts are, as far as possible, the same. Other readers might say that the acts are in fact equally right or wrong, but the people are not – to

think otherwise is to confuse character evaluations with act evaluations. Some might say that the acts themselves are far too different to meaningfully compare, and so it is no surprise that they are morally different. Others might say that the acts are intuitively unequal in terms of rightness or wrongness, but this difference has nothing to do with character - perhaps the moral differences are wholly explainable in terms of occurrent motivation or some other features of the situation that do not amount to character (we will see in the next chapter, the distinction between sporadic moral motivations and those moral motivations inextricably bound up with our moral character will be important).⁵ Working through these differing intuitions and figuring out how we ought to treat such cases is a key goal of this dissertation. My general goal is to argue that in *some* cases, a person's moral character *should* influence our moral evaluations of the rightness or wrongness of their acts. In particular, I won't be trying to provide a complete account of rightness here, but I will argue that adequate theories of rightness should simultaneously meet the following two conditions:

The Consequences Condition: for at least some (but not all) cases, the rightness or wrongness of an act should be determined conclusively by the act's consequences (i.e. how well the act promotes well-being): though there may be competing moral considerations, no competing moral considerations could outweigh the consequences in determining the rightness or wrongness of the act.⁶

⁵ In Chapter 2, I'll describe in some detail this important distinction, which has been developed largely in discussions about virtue ethics (see Hursthouse 1999; Hurka 2006; Cokelet 2015). The main idea is that certain motivations stem from engrained character dispositions while others do not. Importantly, my claim will not be that only character-influenced motivations have moral value, but instead that the ways in which certain motivations are bound up in our moral character - our entrenched and habitual psychological patterns of interpreting, understanding, acting in the world - will inform and help us determine the value of certain acts.

⁶ I intentionally leave the term "consequences" rather open here. There are a great many important distinctions drawn by consequentialists and those interested in

The Character Condition: in some (but not all) cases, the moral character of a person should influence the determination of the rightness or wrongness of that person's acts.⁷

Before I say more about how I'll be interpreting these conditions or how the argument for them will go, it will be helpful to briefly consider what it would mean for our ethical theorizing if we accepted them.

1.2 Implications of the Consequences and Character Conditions

The main project in this dissertation is developing and defending a hybrid account of rightness and wrongness, according to which the rightness of some acts should be *conclusively determined* by their consequences alone (the Consequences Condition) while the rightness of some (other) acts should be *partially influenced* by the character of the agent who performed them (the Character Condition). Before I outline the argument that I'll be providing in support of these two conditions, it is worth briefly taking some time to explore what it would mean for normative ethics if both these conditions held, because it would mean a great deal: in short, that nearly every major ethical theory and many of their most important variations would require at least some revision.

Let's begin with consequentialism. While consequentialist theories of rightness might disagree about exactly which consequences are good, and how that goodness

consequentialism between different ways of cashing out the term. I discuss these in greater detail in Chapter 2.

⁷ Importantly, the term "influence" in the Character Condition is used contrastively to the term "conclusively determine" in the Consequences Condition. That is, the Character Condition does not claim that the rightness of any act is ever wholly determined by the moral character of the person who commits that act. Instead, the claim is that among several potential moral considerations, for *some* (but not all) acts, the moral character of the agent is *one of* the moral considerations that, when taken together, determine the rightness or wrongness of the act.

should be added up, they will agree that it is the consequences alone, counted up somehow that should determine whether an act is right or wrong. This commitment is, of course, perfectly compatible with the Consequences Condition: if the Consequences Condition holds, it does not falsify consequentialism. However, consequentialism is not at all compatible with the Character Condition (as I'll spell it out), since some of the particular ways in which character should influence moral judgment will not be reducible themselves to purely consequentialist considerations. Now, I should emphasize that there *are* ways of spelling out the Character Condition that *are* potentially compatible with some versions of consequentialism. Consider Philip Pettit's version of consequentialism in *The Robust Demands of the Good* (Philip Pettit 2015), according to which we cannot fully capture what we value or cherish about certain actions without considering some of their counterfactual consequences. Pettit argues that for some actions, such as actions of love or honesty, part of what we value or cherish about them is that the person who performed them would also perform them in certain relevant counterfactual circumstances: that the person has a character disposition to perform them. If our theory of rightness is meant to capture the totality of what we value then, it must be (albeit, slightly indirectly) informed and influenced, in some cases, by the character of the person who performed the action. However, this isn't a particularly strong version of the Character Condition: the goodness of the counterfactual consequences of an action is doing most of the work. The arguments I'll develop in favor of the Character Condition will go beyond this: the ways in which (I'll argue that) character should influence our moral judgment about certain actions will not all be neatly reducible (or reducible at all) to actual or counterfactual consequences.

And so, insofar as we find the case for the Character Condition plausible, we should reject a fully consequentialist theory of rightness.

While deontological views come in a great many varieties, we are most immediately interested here in those that stand in contrast to consequentialist views by denying that the rightness or wrongness of an action should be determined only by way of its consequences but instead by its accordance with certain moral rules, principles, or duties. But, there is tension here with both the Consequences and the Character Conditions. First, and most obviously, the Consequences Condition claims that the rightness or wrongness of some acts should be determined *conclusively* by their consequences alone, and is thus absolutely inconsistent with fully anti-consequentialist deontological theories. But further, on the interpretation of the Character Condition I'll defend, the ways in which character should influence rightness are not straightforwardly representable by moral rules or moral principles: they are not codifiable (and, even if they were, their rightness in these cases does not *derive* from their adherence with moral principles). I should however note here, before moving on, a certain affinity with the view of rightness I will develop and the deontological view that Ross spells out in *The Right and the Good* (1930). Both will treat determinations of rightness as a process of weighing various moral considerations against each other, many of which are non-consequentialist: the views will differ in the types of considerations they include and how they are justified.

Virtue ethics, which recognizes the importance of moral character most centrally of the major normative theories, might seem at first glance to be most friendly to the overall project of this dissertation, but in fact there is tension coming from both the

Consequences and the Character Conditions. Typically, virtue ethical accounts of right action treat an act as right just in case it is the virtuous act, or the act that a virtuous person would characteristically do in that situation (e.g. Hursthouse 1999). It seems likely to me that the acts meant to be covered by Consequences Condition would be coextensive with the virtuous acts in those situations: it is hard to see how, in situations in which the differences in the consequences between options is incredibly vast, that a virtuous person would choose the option that, for example, causes vastly more suffering than happiness. But, co-extension is not identity. Vivaly, according to the view I'll develop, in these types of cases, the right options are right (and the wrong options are wrong) *conclusively* because of purely consequentialist considerations: a virtuous person should characteristically perform them *because* of these consequentialist considerations and not the other way around. The tension between virtue ethical theories of rightness and the Character Condition is slighter, and it depends on the degree of implicit agent-neutrality within the virtue ethical theory: that is, the degree to which an act is virtuous regardless of the agent-relative characteristics of the person performing the act. Different virtue ethical theories can vary here, and those that leave less space for differences in behavior, motivation, and composition of character between virtuous agents will clash more strongly with the Character Condition.

1.3 The Argument

The argument for the Consequences and Character Conditions must overcome two initial hurdles: (1) its conclusion is radically revisionary relative to certain views in the literature: to some moral philosophers, it is so counterintuitive that it is simply and obviously false, and (2) in giving such a high level argument, I must ignore a great

many important disagreements, omit a great many important discussions, overlook a great many important debates, and elide a great many important distinctions about rightness and wrongness. So, I'll need to find some way to overcome these first two hurdles before even beginning to make the positive case for the Consequences and Character Conditions.

What we will need first and foremost is a good reason for dislodging our certainty about our most strongly-held intuitions about rightness. Doing so will clear the space for and lift a bit of the burden off of the argument for the Consequences and Character Conditions. But, moral reasoning and ethical theorizing run on our moral intuitions, so we also need the argument to not push us into *complete* distrust of them. The strategy I develop for doing all this will unfold in several steps. First, I deploy a modified version of evolutionary debunking arguments (against moral realism) to dislodge our intuitions about rightness that are based on the belief that there is an objective fact of the matter about rightness (which we know) or the belief that we know certain features of rightness because they are self-evident.

But, in deploying this argument, I appear to dig the hole deeper for myself. The Consequences and Character Conditions are normative claims and need justification *somehow*. But, if my version of the debunking argument succeeds, then justification from moral realism – potentially our best bet for providing normative justification - is not available. And so, I put myself on the hook for an alternative (and, hopefully more plausible) account of moral justification – one that is not susceptible to the debunking argument I give, but still helps achieve the goal of dislodging our *strong* moral-theoretical intuitions without pushing us to discard them altogether. Fortunately for us,

however, the metaethical view I will develop actually provides *further* reason to dislodge our strongest moral-theoretical intuitions without needing to throw away all of them. Also, fortunately, in developing this metaethical view, I won't need to start from scratch: several philosophers have already developed anti-realist and constructivist metaethical views and the view I develop, perspectival naturalism, is a modification of such views. It is "naturalist" in the sense that it is committed to our best scientific / naturalist understanding of the world *and* it does not appeal to non-natural moral truths. It is "perspectival" in that it is interested in far more of our moral minds than our moral beliefs and in far more of our moral views than the propositional claims they are committed to. Perspectival naturalism modifies one popular type of what we might call functionalist anti-realist or constructivist view (according to which moral-theoretical justification derives from how well the moral theory in question facilitates the fulfilment of some narrow function of morality) (Kitcher 2011; see also Mackie 1991's critical discussion of morality as a device for fixing limited sympathies) by treating morality (descriptively *and* normatively) as a *design project* rather than a technology. Thus, according to perspectival naturalism, the best moral theory for creatures like us is the moral theory that is best designed for creatures like us: that best manages the many (sometimes competing, sometimes conflicting) interests of our goals, values, wants, needs, circumstances, psychologies, and tendencies. Thus, moral theories are *constructed* by us, for us. Importantly, perspectival naturalists treat this as a thoroughly unradical claim: this project is *exactly* what humans at all times and in all places have been doing when we engage in ethical discussion and deliberation. And, we are then in the position to make the case for dislodging our most strongly-held moral intuitions

without totally discarding them. We should distrust extremely strongly-held moral-theoretical intuitions because (a) they are *not* justified via knowledge that they are objectively true, (b) they are *not* justified via knowledge that they are self-evident, (c) on the most plausible metaethical view, they descriptively *are* the results of contingent social construction processes, (d) on the most plausible ethical view, they are always subject to change and improvement if a *better-designed* moral theory is available, and (e) on the most plausible metaethical view (or, more accurately, its independently plausible psychological commitments about perspective and our moral psychologies), there can be no *perfect* ethical theory (that is, an ethical theory that perfectly captures all of our moral intuitions in a perspectival framework), and so we must always remain open to the possibility of changes to our held moral theories that we would accept. And so, the burden is slightly lifted – at least enough to justify undertaking the project of arguing for the Consequences and Character Conditions despite the fact that doing so necessitates clashing with or ignoring a great many discussions, debates, and disagreements about rightness and wrongness. But further, this is not done at the cost of forcing us to wholly reject all of our moral and moral-theoretical intuitions: since perspectival naturalism is essentially constructivist, moral justification *must* come ultimately from our moral and moral-theoretical intuitions (of course, via a process of refinement: for perspectival naturalism, a design process). And further still, we now have (in some sense) a novel account of moral theoretical justification that we can put to use in making the case for the Consequences and Character Conditions: we should adopt them if moral theories that do are, *ceteris paribus* better designed for us than those that do not.

The arguments for the Consequences and Character Conditions unfold similarly: we look at the logical alternatives to each of them and show why the Consequences and Character Conditions should be preferred. We compare the Consequences Condition to the Always-Consequences Condition (which states that the consequences should *always* conclusively determine an actions rightness or wrongness) and the Never-Consequences Condition (which states that the consequences should *never* conclusively determine the rightness or wrongness of an action). We then compare the Character Condition to the Only-Character Condition (which states that character is the *only* consideration that should influence our determinations of rightness or wrongness) and the Not-Character Condition (which states that character should not at all influence our determinations of rightness or wrongness) Part of these arguments depend on each other: if the Consequences Condition is independently plausible, then we should reject the Only-Character Condition (because they are inconsistent), and if the Character Condition is independently plausible, then we should reject the Always-Consequences Condition (again, because they are inconsistent). And so, much of the work will be showing that the Consequences Condition is preferable to the Never-Consequences Condition (that is, showing that at least sometimes the consequences *should* conclusively determine the rightness of an act) and that the Character Condition is preferable to the No-Character Condition (that is, that showing that at least sometimes character *should* influence our determination of the rightness of an act). In both cases, the arguments will show that the Consequences and Character Conditions better manage our natural evaluative tendencies and our values and thus, are better designed than these alternatives.

1.4 The Plan

It will be helpful here to briefly outline the structure of the dissertation. In Chapter 2, I develop and describe in detail Consequences and Character Conditions as well as the metaethical view that underlies the project of this dissertation, perspectival naturalism. In Chapter 3, I make the case for perspectival naturalism and develop the ground-clearing or burden-shifting argument that our strongest moral-theoretical intuitions should be dislodged but not altogether abandoned. In Chapter 4, I make the case for the Consequences and Character Conditions. In Chapter 5, I summarize the project, explore the implications for other metaethical and normative ethical theories in much greater detail than I did here, and suggest some directions for further research. Finally, in the appendix, I briefly examine some issues related to truth, relativism, and pluralism for perspectival naturalism.

Chapter 2

Up to this point, I have not spelled out in much detail either the Consequences and Character Conditions or the metaethical view I'll be using, perspectival naturalism. I take up the task of doing so in Chapter 2. I discuss in more detail how I am interpreting the key concepts of *consequences* and *character* as well as the types of cases that I take each condition to cover. Then, I describe and develop key features of the metaethical view that will underpin the argument in the rest of the dissertation.

Chapter 3

The goal of Chapter 3 is to provide a ground-clearing or burden-shifting argument that will justify the relatively revisionary and high-level nature of the argument for the Consequences and Character Condition. That argument will unfold in

several steps. It begins with the development of what I call the “radical debunking argument,” a modification of more familiar evolutionary debunking arguments against moral realism that aims to show that we should not accept (strong versions of) moral realism or straightforward versions of moral intuitionism (according to which we have moral knowledge of certain self-evident truths). I then make the case for perspectival naturalism, the metaethical view that I believe is most plausible, given the radical debunking argument. Finally, I lean on perspectival naturalism to provide several reasons for dislodging our most strongly-held moral-theoretical intuitions without forcing us to abandon them altogether, which will provide us with the ground-clearing and burden-shifting argument we needed.

Chapter 4

In Chapter 4, I make the case for the Consequences and Character Conditions, based on the metaethical view and its account of moral-theoretical justification developed in the previous chapter. Beginning with the Consequences Condition, I contrast it with its logical rivals, which claim that the consequences alone should *always* conclusively determine the rightness of an action or that they *never* should. I provide several reasons to prefer the Consequences Condition over these alternatives. Then, I turn to the case for the Character Condition, which is the bulk of the focus of the chapter. I lay out several arguments in favor of the Character Condition. I then finish the chapter by considering and responding to several objections to the case for the Character Condition.

Chapter 5

In Chapter 5, we will take a step back and look at what I have argued so far. I will provide a summary of the overall argument of the dissertation. Then, I will expand the brief discussion we had earlier in this chapter about the implications of views I've developed, and we will look in much more detail at how perspectival naturalism and the Consequences and the Character Conditions should inform our opinions and evaluations of several extant views in metaethics and normative ethics. Finally, I will suggest several potential directions for future research based on the groundwork that I have laid in this dissertation.

Chapter 2: Overview

2.1 Introduction

In this chapter, I will set out in more detail where we want to go. In particular, I want to describe the two main normative claims that this dissertation aims at defending, the Consequences and Character Conditions, as well as the metaethical view, perspectival naturalism, that will ground them. Then, in the rest of the dissertation, I will defend these normative claims and that metaethical view.

2.2 The Consequences and Character Conditions

A theory of rightness helps us determine which acts are right and wrong. It might do so in many different ways: by providing an analytic definition of rightness, for example, or by simply denying that *anything* actually is right or wrong. In this dissertation, I do not aim to provide any full or complete theory of rightness. Instead, I defend two conditions, the Consequences and Character Conditions, that I think an adequate theory of rightness should meet:

The Consequences Condition: for at least some (but not all) cases, the rightness or wrongness of an act should be determined conclusively by the act's consequences (i.e. how well the act promotes well-being): though there may be competing moral considerations, no competing moral considerations could outweigh the consequences in determining the rightness or wrongness of the act

The Character Condition: in some (but not all) cases, the moral character of a person should influence the determination of the rightness or wrongness of that person's acts.

Let's unpack these. First, note that neither condition claims that our current concepts of rightness or wrongness meet or are constrained by either of the two conditions. For example, the Consequences Condition does not state that, given our *current* concept of rightness, the rightness of an act should be determined conclusively by the act's consequences. Both of these conditions are *revisionary* (though, I will argue, only

relative to certain normative theories, and not necessarily to others or to everyday understandings of rightness). Relatedly, it is important to emphasize that the conditions aim at a theory of rightness – not of goodness or of praiseworthiness.⁸ In this aspect, the Character Condition especially will strike some as highly revisionary. One important conceptual framework in ethics treats rightness just as a consequentialist notion, and claims that other moral considerations matter for the goodness, praiseworthiness, or moral value of the act (see especially Arpaly 2003). We will have to justify positing and then defend such a (relatively) revisionary account. Further, both conditions use the phrase “should be determined” rather than “is,” “just is,” “is constituted by,” or anything of that nature. The conditions do not aim to report genuine properties of rightness as an objective, completely mind-independent feature of the world. That is because I do not think there is any knowable mind-independent fact of the matter about what is right or wrong: what is right and wrong could be determined by us, but it cannot be discovered by us. So, in addition to being relatively revisionary, the conditions are (in a sense) *anti-realist*. However, given that these Conditions do place genuine constraints on how we should construct a theory of rightness, they are not radically anti-realist or nihilist: they are *constructivist*. We will need to describe in more detail this revisionary, anti-realist, constructivist approach, which we’ll do shortly, and to justify it, which we’ll do in the next chapter.

⁸ This is not to say that goodness and praiseworthiness are irrelevant for us or unimportant for developing a theory of rightness. Instead, here I draw attention to the fact that moral considerations that are taken by some philosophers *not to* figure in accounts of rightness (but instead to figure in accounts of goodness or of praiseworthiness) *will* factor into the account of rightness I develop here.

2.2.1 *The Consequences Condition*

We should also take closer looks at each of the two conditions individually. We should notice that Consequences Condition is weaker than a full-fledged consequentialist theory of rightness, which might state that the rightness or wrongness of acts in *all* cases should be or is determined entirely by an act's consequences: there are no other competing moral considerations for rightness. The Consequences Condition allows that there might be other moral considerations for rightness, but that for some acts, the moral weight of the consequences (or typically, some proper subset of the consequences) alone should make it that we do not need to seriously weigh these other moral considerations in our determination of the rightness or wrongness of the act. This is what it means to be determined "conclusively": that no amount of other moral considerations could outweigh the consequences (or some proper subset of them) for our rightness evaluation, though they might otherwise be genuine moral considerations. We should further notice that it is consistent with the Consequences Condition that in some cases, the consequences of an act might not play *any* role in how the rightness of that act should be determined. I do not defend that possibility here and am agnostic about it – I just point out that the Consequences Condition does not rule it out as a possibility. Additionally, it is consistent with the Consequences Condition that the rightness or wrongness of an act can be conclusively determined by its non-consequentialist features, and even that it could be overdetermined by both its consequentialist and non-consequentialist features. Again, I am agnostic about these possibilities – I am just emphasizing that they are not inconsistent with the Consequences Condition as stated. Stepping back, there are two main takeaways from

this elaboration of the Consequences Condition: first, it claims that an adequate theory of rightness should not completely ignore consequences, and second, it leaves open the possibility that an adequate theory of rightness might not be fully consequentialist⁹.

I have not given a general account of what good or bad consequences are. Importantly, I am interpreting the goodness and badness of consequences in terms of welfare. Consequentialist theories typically interpret value along hedonistic, preference-satisfaction, or perfectionist lines: what makes consequences good is that they make people happy, or satisfy their preferences, or contribute to a flourishing life (see D. O. Brink 2006; Sinnott-Armstrong 2015). Welfarist versions of consequentialism (e.g. Sen 1979), at least as I am using them for the Consequences Condition, should be seen as an emendation and loosening of hedonistic treatments of value: the kinds of consequences I have in mind are those consequences that impact people's physical, emotional, and psychological well-being. This treatment does not explicitly deny preference-satisfaction or of perfectionist theories of value – its concerns simply lie elsewhere. It does not treat welfarist goods as the only valuable goods or even the only goods that might play crucial roles in a full theory of rightness. Instead, our goal here is spelling out the claim that sometimes, consequences matter.¹⁰ Now, the Consequences Condition

⁹ One might worry here that as described, the Consequences Condition comes too close to saying something trivial: consequences matter, except when they don't. But, it is important to remember our goals at this stage of the project. Here, I am not developing a decision-procedure for determining rightness, and thus I am not at this stage looking for a method of weighing moral considerations. Instead, my goal here is to argue for the claim that *sometimes* consequences (alone) matter. Since not every moral theory holds this, it is not a trivial claim. The project of weighing various moral considerations in a separate issue than the issue I am here focused on.

¹⁰ This fact is vital to keep in mind at this stage of the argument. The interpretation I here present of the Consequences Condition is not intended to be *exclusive*: I think it is perfectly plausible, and indeed even likely that there are other interpretations of

will likely strike many as more intuitive, or at least, less counterintuitive than the Character Condition. For some acts, the positive or negative weight of the consequences is so incredibly obvious that there simply seems to be no need to even engage in moral deliberation.

Consider Elizabeth Anscombe's criticism of some (consequentialist) views as problematically treating certain moral questions as open:

But if someone really thinks, *in advance*, that it is open to question whether such an action as procuring the judicial execution of the innocent should be quite excluded from consideration—I do not want to argue with him; he shows a corrupt mind. (Anscombe 1958, 14)¹¹

The Consequences Condition should not be taken to go quite this far, so as to claim that for some cases, it is a moral failing to even consider the question of rightness or wrongness because the answer should be manifest, but the sentiment is similar: sometimes, it should just be obvious what is right or wrong¹². As the early Chinese

“consequences” for which the Consequences Condition also holds. If there are, this does not undermine but in fact supports my argument. The purpose of the particular interpretation I provide here is not to provide the *best* or *only* adequate interpretation of the Consequences Condition but instead to provide an interpretation that I think is particularly plausible. Thus, at this stage, I remain agnostic about a great deal of important questions concerning how to interpret consequences (e.g. actual / expected / counterfactual consequences, direct / indirect consequentialism, etc.) since, for many of these distinctions, multiple interpretations could plausibly make true the Consequences Condition. I return to this issue in slightly greater detail in Chapter 5.

¹¹ Of course, it might seem odd to appeal to Anscombe's criticism of consequentialism to support a consequentialist claim. However, the heart of Anscombe's criticism is, on my view, that from within the moral perspective, some questions cannot be open, some theoretical options must be occluded. While she focuses in this passage on views that cannot be available to a *just* person, I claim that certain views are similarly occluded because of their massively bad consequences.

¹² This of course raises the question of whether the Consequences Condition is susceptible to the particular worry that Anscombe discussed: whether it treats as an open question whether the judicial execution of an innocent person could be right (Anscombe 1958, 14). Importantly, though I do think more needs to be said and filled out about this worry, the Consequences Condition does *not* claim that for every case in

consequentialist Mozi was fond of saying, when it comes to some questions, “there are no fools in the world.” What types of cases and questions do I have in mind here? I won’t present here an exhaustive list of the types of cases that the Consequences Condition might cover, but instead mention a few paradigm types of cases. First, I have in mind for the Consequences Condition highly counterfactual, science fiction type cases of the following sort: an evil scientist (whom you have every reason to believe is telling the truth) gives you a choice between hitting a red button, which will cause every sentient creature in the universe to feel the maximum amount of suffering neurologically possible for that creature for a million years, and then die, or hit the green button, which will do nothing (if you refuse to choose, the scientist will hit the red button). The consequences alone determine what one should do in this situation, and rather trivially so. The second type of paradigmatic case for the Consequences Condition is non-counterfactual, but similar to the first in that the immense *scope* of the consequences is what should conclusively determine the action’s rightness or wrongness: the sheer and massive number of people helped or hurt by the action makes it the case that other moral considerations (including the intentions and motivations of the agents performing the actions) simply could not outweigh the consequences.¹³ The

one option has drastically better (welfarist) consequences than the others that taking that option is morally right. It is weaker than that: it says that in some circumstances (when there are not mitigating factors), the consequences conclusively determine the rightness of the action. Injustice may be just such a mitigating factor – here I do not develop a full account of rightness that would explain all of these mitigating factors and their relations.¹³ On the positive side, think of cases like Jonas Salk’s decision not to patent the polio vaccine, or Stanislav Petrov’s (an air defense officer in the Soviet Union) decision to disobey orders and not order a massive retaliatory nuclear strike against the United States after the Soviet nuclear missile launch detection system he oversaw gave a false alarm on September 26, 1983. On the negative side, think here of cases like the communist purges in the mid-20th century throughout central and eastern Europe,

third type of paradigmatic case for the Consequences Condition is based not only the scope of the consequences but on their magnitude or severity: think here of actions that seem right or wrong not based (only) on the number of people they affect but instead based on the tremendous and staggering degree to which they help or hurt someone.¹⁴

2.2.2 *The Character Condition*

The Character Condition is, I take it, the more controversial of the two conditions, and we will be paying much more attention to it, both here where I will elaborate on what it means and is committed to, and in Chapter 4 when I make the case for it. The Character Condition states that in some (but not all) cases, the moral character of a person should influence the determination of the rightness or wrongness of that person's acts. The idea here is that sometimes, but not all the time, the character of a person committing an act is a moral consideration that, in part, should help determine our judgments about the rightness or wrongness of the act. Importantly, in

China, North Korea, and Cambodia: even if such purges could somehow succeed in rooting out mindsets that prevented some ideal form of society and the perpetrators of these purges actually were acting on these putatively loftier ideals, this should not in any way affect the judgment that these purges were wrong.. Similarly, consider Kipling-style Imperialism, putatively aimed at somehow bettering the lives of the conquered "...new-caught, sullen peoples, Half devil and half child." Even if such imperialists genuinely believed that in conquering and ruling foreign countries while replacing the conquered culture with the imperialist's own genuinely would improve the well-being of peoples in those countries (as opposed to say, providing a thinly veneered and highly racialized excuse for exploitation, enslavement and resource extraction), the badness of the consequences of their actions makes it the case that when determining the wrongness of their actions, their intentions are (intuitively, or not terribly counter-intuitively) beside the point.

¹⁴Think here of heinous and vile crimes such as those committed by the Glanton Gang in Northern Mexico and the American Southwest and depicted in Cormac McCarthy's *Blood Meridian*, of the torture and killing of Arian Najibpoor (Blumenfeld 2003, 86) or Junko Furuta, or of the Japanese treatment of "comfort women" in East and Southeast Asia before and during World War II. Though some of these examples include both scope and magnitude, the magnitude alone would conclusively determine their wrongness, even if their scope was limited to just one person.

cases where moral character should influence our judgments of the rightness of an action, I do not claim that character *alone* should do so. Other moral considerations apply – hence I use the term “influence” rather than the term “conclusively determined.” Further, it follows trivially from the Consequences Condition that character should not be a determining or even serious moral consideration for *all* acts, and this is an important point. Thus, moral character should not play a role in *all* act evaluations, and it should not play the *only* role in the act evaluations where it does matter. But, the most immediate question concerning the Character Condition is how to interpret “character.”

I won't here attempt to provide a full account of what I mean by character, for two main reasons.¹⁵ First, the argument I'll be giving for the Character Condition does not depend on a complete account of what character is, but instead on particular aspects of character. Second, and more importantly, I won't be using “character” in any revisionary or particularly atypical ways: I mean by “moral character” what people in general mean by it. What people in general mean by moral character is captured fairly well in the treatment of moral character by supporters of person-centered accounts of moral judgment from empirical moral psychology, which we will look at in more detail

¹⁵ For us, *character* is a concept that models parts of our psychologies. Since the parts of our psychologies that we are modeling when we talk about character are not a purely natural kind, character is not a fully objective, completely mind-independent part of reality, though it is grounded in (and, at least supervenes on) the real parts of a person's psychology that we are modeling when we talk about when we use characterological terms exist in the world. In other words, since we use *character* as a concept to demarcate chunks of our psychologies that we take to be especially morally important, and since *morality* itself is a constructed concept, *character* is as well.¹⁵ This means that (especially, given the metaethical view I'll describe and defend shortly) there should be no fixed and permanently stable definition of “character.”

in Chapter 4 (Pizarro and Tannebaum 2012; Uhlmann, Pizarro, and Diermeier 2015): In addition to dispositional traits, such as honesty, kindness, or courage which have been a locus of attention for moral philosophers (see, for example C. Miller 2013), moral character on person-centered accounts also includes “global” character evaluations such as whether a person is a “good” or “bad” person as well as a person’s deeply-held moral perspectives and commitments, principles, and values (Uhlmann, Pizarro, and Diermeier 2015, 74; Pizarro and Tannebaum 2012, 92–93).

What are the paradigm cases in which our evaluation of a person’s character should influence the evaluation of their acts? Again, I will not provide an exhaustive list here, but mention some of those that will be especially important in this dissertation. I should emphasize here that these paradigm cases likely will not be and are not expected to be intuitive to everyone: some might find them immediately plausible while others will not. The arguments I will develop for the Character Condition do not depend on its immediate intuitiveness. The first way that character evaluations might inform act evaluations is when an act is particularly informative about a person’s character: actions that genuinely and strongly indicate positive or negative aspects about a person’s moral character should be evaluated (*ceteris paribus*) as relatively more right or wrong. The second way that character should influence some of our act evaluations, which has recently been explored and defended by Phillip Pettit (2015), occurs when we can only fully specify what we value or cherish about an act by including features of the character of the person who did it. We begin with the idea that we can distinguish between acts that share similar or identical immediate consequences but which we interpret very differently. Consider two examples: the distinction between telling the

truth and acting out of honesty, and the distinction between helping someone and acting out of affection or love for them. We can imagine particular cases in which the act and its immediate consequences are the same for each of these distinctions, but what we value or cherish about the acts of honesty or affection goes over and above the immediate benefits they provide: we don't value or cherish only the fact that the person provided some immediate benefit in a particular situation, but that they are disposed to do so across a variety of circumstances. If it is further the case that sometimes we really ought to act out of, say, honesty or affection (in these given senses), rather than merely providing the thin benefit of, say, truth-telling or helping, then a theory of rightness will need to capture this.¹⁶ The third way that character evaluations should influence act evaluations is if an act was performed out of the deepest and most genuinely-held moral commitments and perspectives of a particular person. Now, the claim here is not that *any* act is made right if it is performed on the basis of a person's deepest-held moral commitments (this would be inconsistent with the Consequences Condition). Instead, the claim is that if an act is in some sense minimally morally acceptable but is performed out of deeply-held moral commitments or perspectives (that are themselves morally acceptable), it should be treated as relatively more right than it otherwise would be.¹⁷ The fourth way is if the act was performed out of *personal and particular* but

¹⁶ One might worry here that the person's motivations, rather than their character is influencing the rightness or wrongness of the act. Importantly, for some of these cases, the influence from character is indirect: that is, we cannot fully specify the goodness of the motivations of the agent without taking into account of the moral character of the agent, and thus (in these cases) it is through these character-grounded motivations that moral character should influence our judgments of rightness and wrongness.

¹⁷ Again, the idea here is that moral character influences rightness because of the tight connection between a person's deeply-held moral commitments and perspective and their moral character.

morally valuable deeply-held commitments or perspectives. This is subtly different from the previous case: in that case, the added rightness came from our appreciation of the person's integrity, but in the latter case, the added rightness comes from our appreciation of the person as a particular, unique individual.

It is imperative here to be clear on the role of motivation and commitment in the types of cases I have in mind. After all, for many of these cases, one might respond that we can capture all our intuitions that features of the agent matter for evaluating the rightness or wrongness of the act beyond its consequences can be captured without reference to moral character but instead by reference to the agent's motivations and commitments when they performed the action. But, not all motivations (and even, not all moral motivations) are equal. Consider the following cases described by Karen Stohr:

Imagine four wealthy people who write a check to a worthy charity for the same, substantial amount of money. Minerva writes the check because she cares about the people served the charity and wants to make their lives better. Albus writes the check because he believes it's the generous thing to do and he wants to do what is generous. Gilderoy writes the check because he wants his name to appear on the charity's annual published list of benefactors. Petunia writes the check because she wants to spite her sister, who gave her the money for the purpose of paying off her credit card debt. (Stohr 2018, 453)

Virtue ethicists since Aristotle have distinguished good from bad moral motivations and further distinguished sporadic or intermittent moral motivations from moral motivations that stem from our character (Aristotle 2014; Hursthouse 1999; see also Sarkissian 2010): some of our moral motivations, commitments, and perspectives are deeply and inextricably linked to our moral characters, and it is those motivations, commitments, and perspectives that will be most of interest to us in our discussions of the Character Condition. There are different ways of cashing out and justifying this distinction.

Rosalind Hursthouse does so by treating “moral motivation” (i.e. doing something because you believe it is morally right) as essentially tied to our moral character:

... ‘because she thought it was right’ (‘from (a sense of) duty’, etc.) is an ascription that goes far beyond the moment of action. It is not merely, as grammatically it may appear to be, a claim about how things are with the agent and her reasons at that moment. It is also a substantial claim about the future (with respect to reliability) and, most importantly, a claim about what sort of person the agent is—a claim that goes ‘all the way down’. (Hursthouse 1999, 95)

Thus, importantly for Hursthouse, moral motivation comes in degrees: insofar as our moral character is underdeveloped, so too must be our moral motivation. Hursthouse imagines small children who, after (for example, helping someone who was hurt) saying “It was the right thing to do,” and argues that:

...we still do not suppose that they are acting because they think (i.e. believe) that what they are doing is right, or their duty, or required by principle... They are simply too young to have those beliefs. They do not, and counterfactually would not, say and do a whole range of the other things required for the ascription of such beliefs.... By and large, their moral beliefs, or their values, are still not entirely their own, but those of their mentors or their peer group. (Hursthouse 1999, 110)

That is, for Hursthouse, for a motivation to be moral (in the sense she is interested in), it must be tied in a deep way to the person who has it: the person must understand and own the motivation, it must be entrenched in her psychology.

Now, we do need to be careful here: Hursthouse is focused on the relation between moral motivation and *virtuous* moral character while I am interested in moral character in general and (here) the motivations that stem from it, whether that character is morally good, bad, virtuous, vicious, or neutral. Another important difference is that, unlike Hursthouse, the account I develop is not grounded in virtue. This will help us avoid a criticism of certain virtue theories that goes like this (Hurka 2006; Markovits

2010; Crisp 2015; see Cokelet 2015): certain seemingly obvious virtuous acts and motivations (e.g. a soldier who wasn't particularly courageous in general jumps on a grenade to save others), are for some, intuitively just not made good by any dispositional character trait (and so, we might wonder whether virtuous dispositions alone can fully ground an ethical theory). Now, there may be good responses here available for virtue ethicists, but I want to emphasize that the claim of the Character Condition is only that *some* acts are made good (or more good) or bad (or more bad) because of their connection with dispositional character traits, and so these types of counterexamples will not be a major concern for us.¹⁸

There are empirically-informed stories about how certain motivations and commitments can be tied to dispositions. One way of doing so is to utilize the CAPS model from social and personality psychology (Mischel 1999; for examples of philosophers utilizing CAPS, see: Kamtekar 2004; Russell 2009; Snow 2009). According to this model, there are certain “cognitive-affect units that “tend to be relatively enduring psychological structures in a person’s mind...are sensitive to their own activating conditions, and upon being activated they serve to cause the formation

¹⁸ Since, on my view, character should influence some, but not all act evaluations, the question naturally arises how to distinguish when character should and should not play a role. I don't (and given the metaethical view I defend in this dissertation, don't aim to) give a full answer to this question here, but I have suggested the main and most likely ways in which I think character should influence some act evaluations: (1) when an act is particularly informative about a person's moral character, (2) when the goodness of the act / its motivations cannot be fully specified without reference to the person's moral character, (3) when the act is performed out of certain deeply-held moral commitments and perspectives, (4) when the act is performed out of certain uniquely-held moral commitments and perspectives. I make the cases for these in Chapter 4. I leave open the possibility that there may be other circumstances in which moral character should influence certain moral judgments, but it would need to be justified in the same way as I justify these four circumstances in Chapter 4: by showing that the best-designed moral theories for us capture them.

of occurrent mental states like beliefs” as well as desires, values, and goals” (C. B. Miller 2016). But, of course, there are many other empirically-informed ways of cashing out this distinction between motivations dependent on character and motivations that aren’t. For example, Hagop Sarkissian has developed an empirically informed account aimed at making sense of Confucius’s audacious claim in *the Analects* 2.4 that, after a lifetime of moral cultivation, his motivations and desires effortlessly conformed to what was right:

We now have plausible reason to believe Confucius's claim to effortless, spontaneous, and appropriate response. Using Damasio's terminology, Confucius's program of incessant self-cultivation will produce countless somatic-markers, extending the range of stimuli that become automatically marked. The accrual of these markers over time fine-tunes and accelerates the decision-making process; at the limit, the correct course of action would come to mind immediately, with compelling emotional valence. (Sarkissian 2010, 7)

That is, through familiar processes of habituation and training, certain emotional responses will tend to fire off more frequently in certain situations, and these emotions in turn will predictably produce certain motivations to act in certain ways. Not all motivations (in fact, very few motivations) are produced by such a process: some are sporadic and have little to do with our entrenched ways of being, thinking, understanding, and perceiving the world. But those motivations that *do* have a great deal to do with our entrenched psychological dispositions will be of interest to us as we defend the Character Condition.¹⁹

¹⁹ It is important to mention here one small but important area of divergence from typical Aristotelian-inspired accounts of character and the notion of character I am interested in here. I include in my notion of character a person’s “deeply-held” commitments, perspectives, and values but “deeply-held” here does not simply mean “dispositional.” Instead, it means the commitments, perspectives, and values that a person most strongly identifies with and accepts. Importantly, this might not always have a *strong* causal or correlative relationship with their behavior: they might fail to

2.3 Perspectival Naturalism

But before we can lay out the arguments to support the Consequences and Character Conditions, we need to clear some space. Philosophers have diverse and strong commitments, of varying degrees of specificity, about rightness and wrongness. Why should a philosopher for whom it is highly intuitive that rightness is purely consequentialist, for example, even begin to consider the Character Condition? Insofar as the Consequences and Character Conditions are radically revisionary relative to nearly every extant theory of rightness (though, not necessarily to all or commonsense theories of rightness), for any argument in support of the Consequences and Character Conditions to even *begin* to be convincing we must dislodge some of these firmly-held commitments. And yet, we do not want to go too far: our arguments aim to be revisionist about some of our moral-theoretical intuitions, not eliminativist about them. I am not aiming for some radically skeptical claim that our intuitions about rightness (or our intuitions about morality more generally) are completely and hopelessly off-track or useless, but instead for the claim that we should *always* be open to the possibility that they require revision. To do this, we will take a detour through metaethics, and I will argue in Chapter 3 in support of a metaethical view, perspectival naturalism, that will justify the revisionary aspects of this project without forcing us to completely abandon our moral intuitions. But, before we set out that argument, I want to describe in more detail what perspectival naturalism is and how it compares to rival metaethical views.

live up to their commitments sometimes. This is not to say there is *no* connection to behavior: we might rightly question whether someone genuinely holds a commitment if they *never* act on it.

The central commitment of perspectival naturalism is, in a nutshell, that the best moral systems for us are the moral systems that are best *designed* for us. Before we start discussing some of the metaethical commitments of perspectival naturalism, let's briefly unpack this statement. First, we should note that the key term here is "design." Perspectival naturalism is inspired by and shares a great deal in common with anti-realist (see especially Mackie 1991; Kitcher 2011; Flanagan, Sarkissian, and Wong 2016) and constructivist views (Street 2010; Copp 2007) that treat morality (both descriptively and normatively) as an *invention* or *technology* that humans created out of our needs, circumstances, goals, and values. Perspectival naturalism modifies these views, however, by suggesting that instead of understanding morality (again, both descriptively and normatively) as an invention or as a technology, we should understand it as a *design* project, akin to architectural, interior, systems, industrial, or user experience design. Practically, this will mean that instead of thinking of morality as having a single or primary unifying function (as many inventions, such as hammers or cars do) that will help determine how well justified different moral theories and claims are (i.e. how well they serve that single or primary function), we instead think of morality as a project of bringing into alignment a great many considerations, influenced by our goals and needs, our values, our circumstances and our abilities. In Chapter 3, we'll explore in much more detail this design paradigm and what it will mean for morality. The second thing we should note is that the central commitment of perspectival naturalism is framed in terms of which moral systems are *best*, not which are true. This is for two main reasons: first, perspectival naturalism (as the word "perspectival" in its name suggests) is interested in much more of our psychology than

the bits that contain truth-apt beliefs, and so it aims to evaluate moral systems not *only* in terms of the truth of the moral claims they make. Second, while perspectival naturalism is far from unconcerned about which moral claims are true, it pays much more attention to justification than truth. In particular, I should emphasize immediately here that perspectival naturalism does *not* claim that the best moral systems are those comprised of all and only objective or mind-independent moral truths. I explore some issues related to truth, relativism, and pluralism for perspectival naturalism in the Appendix of this dissertation. Third, we should avoid a potential misunderstanding here about the expression “best...for us.” When I use the expression “best moral systems for us,” the “best...for us” should not be interpreted to mean “the most beneficial for us.” Certainly, adhering to the best moral systems will provide some benefits to us, but this is not what justifies them according to perspectival naturalism. In fact, it is perfectly possible, and indeed almost certainly the case that the best moral systems will, in particular circumstances, suggest actions that are less beneficial to a particular person than other available options. Instead, the “best...for us” should be taken to mean “best for *creatures like us*, with the psychologies, values, goals, needs, desires, perspectives, abilities, and circumstances we have or find ourselves in.” Importantly, coming to understand what is best for us in this broad sense is *part of* the moral design project.

We should describe quickly, before we discuss the view in more detail, what perspectival naturalism has to say about the key components of a metaethical theory: specifically, its commitments concerning moral ontology, moral epistemology, and moral semantics. Concerning moral ontology, perspectival naturalism is *apathetic* (rather than agnostic) about the existence of fully mind-independent, objective moral

features of the world. It is not committed one way or the other to their existence. The reason for this actually derives from the moral epistemological commitments of the view: namely, that we could not know fully mind-independent, objective moral truths, even if they did exist. I should mention here that by “objective moral truths” I intend a rather strong sense of objectivity, which will amount to fully mind-independent truth. This is not the only way “objectivity” can be used here (it is at odds with, for example Copp’s (2007) use of the term), but my reasons for doing so will become more clear in Section 2.23. Finally, perspectival naturalism is a constructivist view that takes very seriously the idea that moral systems are social constructs, and this has important implications for its treatment of moral semantics.²⁰ In particular, for perspectival naturalism, there is little if any fixed meaning of moral terms or claims. Whether moral claims report truth-apt propositional attitudes or express non-truth apt emotions, for example, depends on how the relevant moral systems are constructed.

So, we have seen what perspectival naturalism is, in a nutshell, and what its most general metaethical commitments are. Now, we are in a position to describe and

²⁰ I should say more about what I mean by “social construct” since the term is easy to misinterpret. By “social construct” in general, I do not mean only a *purely* social construct, that is a construct completely untethered from nature and natural kinds. On my view (and I think any sensible view of social constructs) some social constructs could be partly grounded in and even supervene on natural kinds (e.g. genes, species, catalyst, jade, etc.). Of course, not all social constructs are merely social kinds (for example, baseball and ballet), and morality is this more robust type of social construct. Finally, one might wonder if treating morality as a social construct implies, perhaps surprisingly, that there is no morality outside of society. It does just that, though I am using a very loose sense of the term “society” and referring specifically to human morality. The reason for this is straightforward: our moral systems are shaped by our moral psychologies, and our moral psychologies were shaped by the evolutionary pressures of living and surviving in small-kin groups. Even if you take a human out of society, you cannot take society out of a human – it’s embedded deep in our moral psychologies.

develop the view and its commitments in more detail. Perspectival naturalism is motivated by three main insights: (1) morality is and should be what we have actually been doing, creating moral systems based on our values and goals and constrained by our circumstances and moral psychologies, (2) Understanding morality in this way opens up new considerations for evaluating moral systems, and (3) The treatment of our moral minds should be in terms of perspectives rather than sets of propositional attitudes; the treatment of our moral theories should be in terms of (paradigm) systems rather than in the sets of claims they are committed to. Perspectival naturalism has five main commitments:

1. **Naturalism / Descriptive adequacy:** Moral theorizing should be informed by our best descriptive understanding of the world; this understanding will be naturalistic
2. **Anti-realism:** There are no *knowable* objective, fully mind-independent moral truths
3. **Non-Nihilism / Constructivism:** Anti-realism does not entail the abandonment of morality; moral systems are ultimately justified not by being objectively true but via our moral standpoints
4. **Perspectivalism:** An adequate model of the mind must treat the mind as perspectival; perspectives do not reduce without loss to sets of beliefs
5. **The Design Paradigm:** The best moral systems for us are the moral systems that are best designed for us

Let's look at these in more detail.

2.3.1 Naturalism / Descriptive adequacy

“Naturalism” is not the clearest term in all of philosophy. The term itself was developed largely in a metaphysical and epistemological context to pick out a philosophical worldview that aligned itself with science in trying to understand the world and human beings and that posited nothing “supernatural” (Kim 2003; Papineau 2015). Of course, there is an immediate concern here that what is seen as supernatural in one worldview is often enough seen as perfectly natural in other worldviews (e.g. *qi*

energy in some east Asian philosophical, medical, and religious traditions, karma in certain south Asian traditions), and it seems difficult to distinguish between seemingly naturalist entities that do not exist like phlogiston or aether or pre-Newtonian occult properties and putatively supernatural entities like *qi* or karma. In the metaethical context, “naturalism” is used mainly as a contrast to non-naturalist views, according to which moral truths are self-evident “non-natural” facts like the facts of mathematics.

Perspectival naturalism’s commitment to naturalism should be interpreted primarily as a commitment to descriptive or empirical adequacy: insofar as our metaethical or normative theorizing relies on or interacts with descriptive or empirical claims, we should ensure that we are relying on only the best justified descriptive and empirical claims. This commitment to naturalism will have three general upshots for us. First, perspectival naturalists take evolution to be a key explanatory factor for understanding our moral minds. Second, perspectival naturalism does not seek any moral justification or information from supernatural sources (e.g. the laws of God) or non-natural sources. Third, perspectival naturalists treat models of the moral mind whose focus is aimed entirely on propositional attitudes as empirically inadequate. These upshots will inform and motivate the other main commitments of perspectival naturalism.

It is important to note that I will leave this commitment to naturalism largely undefended here. The main reason for this is that there is little if anything I could contribute here that would sway those who hold entrenched views about naturalism one way or the other. Though it is by no way a total defense of naturalism, I can provide here a *pro tanto* reason for at least not being terribly bothered by perspectival

naturalism's commitment to naturalism: if my arguments succeed, there are few if any *moral* reasons to reject perspectival naturalism: it will practically lead to many of the same moral conclusions as other moral theories would.

2.3.2 *Anti-realism*

There is not universal agreement about what exactly moral realism is committed to. We will not here examine in detail the varieties of accounts of moral realism, but I do want to briefly focus on two issues. First, many accounts of moral realism take moral realism to be committed to the claim that there are *objective* moral truths (Shafer-Landau 2005). But, it is not clear how we should interpret "objective" – as Lorraine Daston has convincingly argued, the term is incredibly messy.²¹ In particular, it is no easy challenge to determine just how mind-independent we think objective truths should be. Second, even if we did agree on a particular account of objectivity for moral realism, it might be the case that objectivity is beside the point. For example Copp (2007) has argued that we should count a metaethical view as realist so long as it can generate moral truths – whether they are fully objective and mind-independent or not.

²¹ She writes ["Objectivity"] refers at once to metaphysics, to methods, and to morals. We slide effortlessly from statements about the 'objective truth' of a scientific claim, to those about the 'objective procedures' that guarantee a finding, to those about the 'objective manner' that qualifies a researcher. Current usage allows us to apply the word as an approximate synonym for the empirical (or, more narrowly, the factual); for the scientific, in the sense of public, empirically reliable knowledge; for impartiality-onto-self-effacement and the cold-blooded restraint of the emotions; for the rational, in the sense of compelling assent from all rational minds, be they lodged in human, Martian, or angelic bodies; and for the 'really real', that is to say, objects in themselves independent of all minds except, perhaps, that of God. In its thick layering of oddly matched meanings - it is not self-evident, for example, what the repression of the emotions has to do with the ontological bedrock - our concept of objectivity betrays signs of a complicated and contingent history, much as the layering of potsherds, marble ruins, and rusted cars would bespeak the same in an archeological site" (1992, 597).

This would count some versions of metaethical constructivism, including perspectival naturalism, as realist. I won't try to definitively settle any issues about what the best account of objectivity and thus moral realism is other than to say this: there is an interesting, historically and philosophically important category of moral theories for which the truth and justification of moral claims does *not* ultimately derive from features of *our* minds and perspectives. This category includes, for example, Plato's theory of forms according to which moral claims derive their truth from correct correspondence to a world of forms that exists independently of us; divine command theories according to which moral truth and justification derives from God's commands, certain natural law and naturalist theories such as those of Aquinas and possibly Mencius according to which the truth of moral claims derives ultimately from our natures, given to us by God or Heaven; and intuitionist theories according to which moral claims are self-evidently true.²² I'll simply stipulate that for our purposes, theories that belong to this category are realist while theories that do not are anti-realist. Thus, I treat metaethical theories according to which there are mind-*dependent* true or justified moral claims as anti-realist. If this seems too high a bar, one is welcome to read "strongly realist" or "strongly objective" when I say "realist" or "objective."

2.3.3 Non-Nihilism / Constructivism

Take nihilism to be the claim that moral anti-realism entails that no particular (non-nihilistic) moral or normative claim could be justified or true. Perspectival

²² Mind-independence here should not be taken to mean that minds play no role whatsoever in morality – for example we need minds to grasp self-evident truths. The claim instead is that is that these moral truths are not *made true* by features of our mind alone. For example, self-evident moral truths would be self-evident for anyone who could grasp the relevant concepts – they are made true by features of the concepts, not by our minds.

naturalism, like many other anti-realist views, rejects nihilism. Perspectival naturalism derives this rejection from its constructivist ancestry: since what ultimately justifies or makes true moral and normative claims is our values, and since we have values, there can be justified and true moral and normative claims. In particular, perspectival naturalism takes its cues here from what Sharon Street has called “Humean constructivism” (2010) According to Street, all constructivist metaethical views share the idea that moral claims are true because they are somehow entailed from within a certain practical perspective that we are able to adopt. Humean constructivism, in particular, is further committed to the claim that there are no substantive moral truths that are entailed in *all* practical standpoints.

2.3.4 Perspectivalism

Many metaethical views are only or primarily interested in very small section of our minds: for example, our moral beliefs in the case of moral realism, or our non-cognitive moral attitudes in the case of moral expressivism. A further, seemingly small but in fact vitally important commitment of perspectival naturalism is a commitment to treating our minds as perspectival. Recently, several philosophers have begun to develop accounts of our minds in terms of perspective and have explicitly argued that perspectives are not reducible without loss to sets of beliefs (see Camp 2006; Riggs 2016; Elgin 2010). As we’ll see in the next chapter, perspectives include, in addition to beliefs, values, desires (and other propositional attitudes): patterns of association, what we perceive as having certain properties, which inferences we are likely to draw, where we look for / what we see as evidence, how we interpret new propositions, and which features of things are more prominent or salient or central.

There are two immediate implications of this perspectival view. First, and most obviously, it matters for how our moral theories treat our minds. Our moral goals will include not only having the best moral beliefs, but also having the best moral perspectives. But, second, it also matters for how we think of moral theories themselves. Thus, I will typically use the term “moral systems” rather than “moral theories” when discussing perspectival naturalism. This is not meant to be a radically revisionary move, but instead to be a slightly improved understanding of what we already do. Traditional moral theories, such as utilitarianism, do not consist in and cannot be reduced to merely a set of propositions. They include characteristic perspectives or paradigms, ways of seeing the world, where certain information and considerations are more or less salient, more or less central, more or less occluded. Thus, we are not only interested in justifying moral beliefs but moral perspectives and we are not only interested in justifying moral theories but moral systems.

2.3.5 The Design Paradigm

As we have seen, several anti-realist metaethical views treat morality as an invented technology (Mackie 1991; Kitcher 2011, 2012; Flanagan, Sarkissian, and Wong 2016), often with some particular narrowly-defined function that grounds moral justification: moral theories are better when they fulfil that function better. For example, Kitcher (2011) argues that the main function of morality is to increase the benefits of psychological altruism and decrease the damage of altruism failures, and thus moral theories that better fulfill this function should be preferred. For a perspectival naturalist, this line of thinking is very close to being correct. It fits well with the first three commitments of perspectival naturalism: descriptive adequacy / naturalism, anti-

realism, and non-nihilism. However, perspectival naturalism treats these “moral technology” views as overly narrow (we will see the argument in the next chapter): morality cannot and should not be reduced to having a single primary function. Instead, perspectival naturalism modifies these views, and treats morality not as an invented technology with a narrow function but instead as a project of design. Think here of a landscape architect re-designing a small section of a university campus. They have to manage and align a great deal of connecting and interfacing considerations at once. They need to ensure that the paths follow routes that people will actually use, so as to avoid creating new dirt paths. They need to ensure that the area is accessible for people with physical disabilities. They need to design the area so that it is aesthetically pleasing, and so that it is safe. They need to place objects so that they will not interfere with gas lines or power lines. They need to make sure that the required construction vehicles can actually access the area to build what needs to be built, and they need to do all of this while meeting financial and time constraints. In short, the key idea of the design perspective is just this: that the project of morality, both as it actually has been and as it should be, is like this design project: they are both constrained, shaped, and guided by our goals, values, abilities, circumstances, limitations, wants, needs, and principles. Thus, what a moral system aims at doing is managing and aligning all of these competing considerations in a way that is itself (recursively) aligned with them. Now, this rejection of a single primary function of morality does not mean that anything goes: that no moral values have priority over any other when it comes to how we should align moral considerations. We should instead think of the picture here somewhat similarly to Ross’s moral theory in *The Right and the Good* (1930): some moral

considerations might typically be more important than others, but they all matter and there is no immutable hierarchy of them. For example, on Ross's view, there are moral considerations of beneficence and non-maleficence, of self-improvement and of gratitude and of reparation: they each matter in different ways and what is right to do can and should be informed by many, or even all of them at once (Ross 1930, 21). The design paradigm for perspectival naturalism works much the same way, except that it includes even *more* considerations than our most important moral values: in particular, it will include *design* considerations.

How might this work in practice? Consider just one example here (we'll explore more in Chapter 3). In ancient Greece, an especially central virtue was *xenia*, or "guest-friendship" – a particular type of hospitality aimed at outsiders. Zeus was sometimes called Zeus Xenios, and Homer's *Odyssey* can be seen in part as a literary examination of the virtue (Wilson, 2017). As Emily Wilson describes the concept:

Xenia acquired an extra importance in the era when Greek men were expanding their world. Travelers, in an era before money, hotels, or public transportation, had to rely on the munificence of strangers to find food and lodging and aid with their onward journey. (Wilson, 2017, 21)

Now, the general value or virtue of being kind and hospitable to strangers or outsiders is far from unique to the ancient Greeks and certainly a good value to have across cultures. But, what we see here is that *xenia* was especially *useful* in the ancient Greek context: the economic systems, and thus people's well-being, depended on dangerous sea-faring trade. In addition to being a moral value, *xenia* also served as insurance policy that helped improve people's well-being and peace of mind. They could be assured that if their ship crashed on a distant island or near a distant city (which was a very real possibility) that they would not be treated as enemies. Perspectival naturalism, and

particularly its design perspective takes these additional considerations seriously: it claims that in those circumstances, it really ought to be the case that *xenia* had in especially central role in the moral system. Importantly, the argument here is not that these additional circumstances alone justify the inclusion of *xenia* as a moral value or virtue – it is a good value or virtue in any (or nearly any) circumstances. Instead, the idea is that there are many different moral systems, with many ways of weighing our various values: while some metaethical views might treat the added benefits of *xenia* under the specific circumstances of ancient Greece as irrelevant or beside the point, perspectival naturalism treats them as relevant and in need of consideration.

2.4 Summary

In this chapter, we have taken a closer look at where we want to get to and set out in more detail the Consequences and Character Conditions on theories of rightness as well as the metaethical view which will support them. But, we haven't yet begun to make the case for the Consequences and Character Conditions, or for perspectival naturalism. We will begin to do that in the next chapter as we justify perspectival naturalism on the basis of the radical evolutionary debunking argument.

Chapter 3: The Case for Perspectival Naturalism

3.1 Introduction

In this dissertation, I argue that an adequate account of rightness should meet at least the Consequences Condition and the Character Condition. The Consequences Condition states that for at least some (but not all) cases, the rightness or wrongness of an act should be determined conclusively by the act's consequences (i.e. how well the act promotes well-being): though there may be competing moral considerations, no competing moral considerations could outweigh the consequences in determining the rightness or wrongness of the act. The Character Condition states that in some (but not all) cases, the moral character of a person should influence the determination of the rightness or wrongness of that person's acts. But, as I pointed out in the previous chapter, in making this revisionary (and ambitiously revisionary) claim, I must necessarily elide heaps of important historical and contemporary discussions, debates, distinctions, theories, views, accounts, puzzles, and intuitions about rightness. I need to earn the right to undertake such an ambitiously revisionary project and present my arguments in favor of the Consequences and Character Conditions without addressing all of these heaps of issues in the detail they otherwise deserve. To do so, we'll take a detour into metaethics. I'll argue that on the most plausible metaethical view our intuitions about rightness and wrongness *do* support this type of revisionary project because, on this view, moral theories like a theory of rightness are *essentially* revisable. But, in addition to showing that, taking this detour will present us with two extra challenges. First, I'll need to show that this metaethical view *is* more plausible than its rivals. If it is not independently plausible, it will not do much good for us. Second, since

the ultimate goal of the dissertation is providing justification for the normative Consequences and Character conditions, I'll need to show how normative claims can be better or worse justified on this metaethical view (and the arguments I'll provide in the next chapter in favor of the Consequences and Character conditions will need to be consistent with that account of justification).

We briefly discussed this metaethical view, a form of constructivism I've called "perspectival naturalism," last chapter. According to it, the justification or even truth of moral claims ultimately derives from features of our practical points of view – the standpoints of valuing and of normative or evaluative judgment – but no particular, substantive moral claims can be derived from all practical points of view (Street 2010; Arruda 2017). According to perspectival naturalism, theories of rightness, like all moral theories, are constructions aimed at embodying, expressing, and facilitating our values and moral perspectives, and are also always open to revision and improvement. This characterization immediately puts perspectival naturalism at odds with nearly every other major metaethical view. First, since it is committed to the claim that moral truth and justification do not derive solely from fully mind-independent or objective features of the world, it is at odds with all forms of moral realism (Shafer-Landau 2005).²³ This will include naturalistic versions of moral realism that claim that all moral

²³ There is, so far as I know, not strong agreement about what we mean by "objectivity." Lorraine Daston puts the point exceedingly well: ["Objectivity"] refers at once to metaphysics, to methods, and to morals. We slide effortlessly from statements about the 'objective truth' of a scientific claim, to those about the 'objective procedures' that guarantee a finding, to those about the 'objective manner' that qualifies a researcher. Current usage allows us to apply the word as an approximate synonym for the empirical (or, more narrowly, the factual); for the scientific, in the sense of public, empirically reliable knowledge; for impartiality-unto-self-effacement and the cold-blooded restraint of the emotions; for the rational, in the sense of compelling assent from all rational

claims are made true or justified by conspecific features of humans (Aristotle 2014) and non-naturalistic or intuitionist versions of moral realism according to which moral truths are self-evident (Huemer 2005; Stratton-Lake 2014; Shafer-Landau 2005; Sidgwick 1981). But, perspectival naturalism is also inconsistent with several forms of moral anti-realism. It is inconsistent with metaethical nihilism, according to which there are no true or justified first-order moral claims, because according to perspectival naturalism, there can be and *are* justified moral claims. It is also inconsistent with non-cognitivist anti-realist theories in the vein of emotivism (Ayer 1952), according to which there cannot be true moral claims because moral claims are not truth-apt: even if emotivism is descriptively true for some moral-linguistic community at some point in time, according to our version of constructivism, the meaning of moral terms is not fixed but can be revised. This will similarly present problems for versions of fictionalism according to which our moral claims are technically false but useful (Joyce 2007; Mackie 1991): if our moral claims are falsified by an underlying reliance on moral realism, once that reliance is replaced for perspectival naturalism, moral claims are no longer false (at least, in the same way).

minds, be they lodged in human, Martian, or angelic bodies; and for the 'really real', that is to say, objects in themselves independent of all minds except, perhaps, that of God. In its thick layering of oddly matched meanings - it is not self evident, for example, what the repression of the emotions has to do with the ontological bedrock - our concept of objectivity betrays signs of a complicated and contingent history, much as the layering of potsherds, marble ruins, and rusted cars would bespeak the same in an archeological site" (1992, 597). The sense of objectivity I use in this dissertation should be read as "fully objective, totally mind-independent." This is a very strong sense of objectivity for moral realism – those who think that a weaker sense of objectivity should be used should feel free to read "strongly objective" when I say "objective." In a weaker sense of "objective," i.e. of being justified without the undue influence of personal bias, perspectival naturalism does aim to be and can succeed in objectivity.

3.2 Radical Evolutionary Debunking

3.2.1 Traditional Debunking Arguments

In this section, I will present a key argument for the project of this dissertation: the radical evolutionary debunking argument against moral realism. But before I do, I should begin by saying a little bit about traditional debunking arguments to help draw out what is novel about the radical debunking argument. We will begin with very basic versions of debunking arguments and layer on complexity as we go.

Evolutionary debunking arguments and challenges to moral realism have quite an impressive provenance - Charles Darwin might have been one of their earliest proponents. In *Descent of Man*, he wrote:

It may be well first to premise that I do not wish to maintain that any strictly social animal, if its intellectual faculties were to become as active and as highly developed as in man, would acquire exactly the same moral sense as ours...If, for instance, to take an extreme case, men were reared under precisely the same conditions as hive-bees, there can hardly be a doubt that our unmarried females would, like the worker-bees, think it a sacred duty to kill their brothers, and mothers would strive to kill their fertile daughters; and no one would think of interfering. (Darwin 1871, p. 73)

The phrase “were reared” here should be read, given the previous sentence, in a “nature” rather than “nurture” sense: Darwin is claiming that the habits and instincts that particular species evolve would influence that species’ moral sensibilities, that if the members of some species became capable of higher-order moral thought, they would develop a system of morality that encouraged and justified certain behavior towards which they were genetically predisposed. Whatever social structure they were genetically predisposed to adopt would constrain their attitudes toward the range of acceptable social structures. Whatever practices they were genetically predisposed to follow would constrain the range of practices they would believe are morally right. If

they were herbivores, eating meat would not be a live moral option; if they were carnivores, it would. If they evolved to have a high degree of parental investment, like humans, they would be much more morally concerned with the well-being of their offspring than if they evolved to have an extremely low degree of parental investment.

William FitzPatrick and John McDowell have independently argued against drawing this line of reasoning (Fitzpatrick 2008; McDowell 1995). What this argument fails to give adequate attention to, both philosophers have argued, is an important effect of becoming rational: the development of a capability of autonomous moral reflection. As FitzPatrick argues, “If bees became rational, then, there is no reason to suppose that their moral sensibilities would simply reflect their ecology any more than human moral sensibilities simply reflect human ecology...” (Fitzpatrick 2008). Now, I do not think this reply is quite as successful as it might seem. After all, rational reflection is itself a trait deeply influenced, shaped, and constrained by our evolutionary history.²⁴ But the general point is well-taken: insofar as a species has the capacity to *resist* instinctual

²⁴ The insight underlying evolutionary debunking arguments of moral realism can be pushed quite far, and in some cases too far: a much more general skeptical argument was developed by Alvin Plantinga to argue that (among other things) our confidence in the truth of naturalism (scientific / metaphysical naturalism, not moral naturalism) might not be justified (see for example Nagel 2012; Beilby 2002). Interestingly, Nagel attributes this type of skepticism even to Darwin, who wrote in an 1881 letter “But then with me the horrid doubt always arises whether the convictions of man's mind, which has been developed from the mind of the lower animals, are of any value or at all trustworthy” (Darwin 1881). Yet, most proponents of evolutionary challenges to moral realism would be quite unhappy to find themselves in the same boat as skeptics about naturalism or about the validity of our non-moral rational inferences. The more general skeptical argument fails, or at least is much weaker than the moral debunking argument: in the case of our scientific theories, even if our epistemic faculties are generally unreliable, that does not mean they are completely unreliable, and importantly in the scientific or more generally empirical case we have tools such as observation, experimentation, and experiment design based on falsifying hypotheses that are simply not available in the moral case. This key difference plays an important role in the radical debunking argument.

urges or has space in which their decisions and moral opinions are not fully determined by their instincts, their moral opinions can swerve from their evolutionary history and this straightforward debunking argument (that we straightforwardly always do and believe what our evolutionary history determines us to do and believe) cannot succeed.

Consider the following fairly standard version of a debunking argument.²⁵ I take this first formulation directly from Brosnan (2011, 52):

1. “If our moral beliefs are products of a process that fails to track moral facts, then we do not possess moral knowledge.
2. Our moral beliefs are products of evolution by natural selection.
3. Evolution by natural selection is a process that fails to track moral facts.
4. Our moral beliefs are products of a process that fails to track moral facts.
5. Therefore, we do not possess moral knowledge.”

If knowledge is some type of justified true belief, it follows trivially from the conclusion of this version of the argument that our moral beliefs are false (which is Street’s (2006) position) or unjustified (which is Joyce’s position (2007) (or both)).²⁶ This version of the argument is slightly better than the earlier version: while it retains in premise 2 the oversimplified strong connection between our moral beliefs and our evolutionary history, it makes explicit in premises 3 and 4 that the real issue is the non-veridical influences of natural selection on our moral psychologies. Now as we have seen, premise 2 has been attacked on grounds that our moral beliefs can be the products of autonomous moral reflection (FitzPatrick 2008, McDowell 1995). Derek Parfit has

²⁵ This version is, at a very general level, the argument put forward in Ruse and Wilson (1995), Street (2006), and Joyce (2007).

²⁶ Or, it might be that our moral beliefs fail to be knowledge because they fail to meet some fourth condition on knowledge. Of course, figuring out exactly what this fourth condition is has not been successful, and since most treatments revolve more or less around the idea that the justification is not lucky, for the purposes of laying out the argument I’ll just treat this fourth condition as an issue related to justification.

also argued that premise 3 is false, and that knowing moral truths *was* evolutionary advantageous (Street 2006, pp. 125-126). However, the most conspicuous issue with both this and the earlier formulation of the argument is that they treat all moral beliefs as of one class and our moral psychologies as unified entirely by natural selection. Instead, why not think that our moral beliefs are influenced by a combination of evolution, rational reflection, culture, upbringing, and so on?

Street (2006) provides a formulation of the argument, in the form of a “Darwinian Dilemma” that better tracks this feature of our moral judgment. The Darwinian Dilemma for realist theories of morality is this: either there is no relation between moral truth and evolutionary influences on our moral judgments, or there is (Street 2006, 109). Street argues that if we take the first horn of this dilemma, that there is *no* connection between evolution and moral truth, then we are forced into the discomfiting “skeptical result that most of our evaluative judgments are off track due to the distorting pressure of Darwinian forces” (2006, 109). That is, if there is no underlying connection between evolution and moral truth, and our moral judgments are influenced by evolution, then they are extremely unlikely to track moral truth. If we take the other horn of the dilemma and admit that there *is* a relation between evolution and mind-independent moral truths, then we are forced into the unsubstantiated empirical position that natural selection favored ancestors who were better suited at grasping objective moral truths. In this construal of a debunking argument, Street emphasizes that there are other influences on the content of our evaluative judgments than natural selection alone: “some of them were perhaps evolutionary factors other than natural selection – for example, genetic drift, and many other forces were not

evolutionary at all, but rather social, cultural, historical...[and] there is the crucial and *sui generis* influence of rational reflection” (Street 2006, 114). Instead of denying non-natural selection influences on the content of our evaluative judgments, Street takes the weaker but more plausible view that natural selection is one of many factors in shaping our values, but also that it is especially influential (Street 2006, 114). She claims, “it is only reasonable to expect there to have been, over the course of our evolutionary history, relentless selective pressure on the content of our evaluative judgments...in particular...those evaluative judgments which tended to promote reproductive success” (Street 2006, 114). Of course, one immediate problem is that the content of our deliberative and reflective evaluative judgments themselves are ideas. They are not genetically inheritable traits, which means that evolution by natural selection simply does act on them. Thus, Street restricts her claim to only our “more basic evaluative tendencies.” “Basic evaluative tendencies,” according to Street, are roughly “unreflective, non-linguistic, motivational [tendencies] to experience something as ‘called for’ or ‘demanded’ in itself...proto forms of evaluative judgment” (Street 2006, p. 119). The “more basic evaluative tendencies” are just those basic evaluative tendencies which are genetically heritable traits (ibid.). These are contrasted with our “full-fledged evaluative judgments,” which are the end result of our moral deliberation, influenced by evolution, culture, society, history, and rational reflection (ibid). These more basic evaluative tendencies are the means by which the evolutionary challenge is supposed to go through: they massively influence, guide, and constrain our full-fledged deliberative and reflective evaluative judgments.

While this formulation of the argument is helpful in being more nuanced about the role that evolution likely plays in influencing our moral judgments, it could be cleaner – we don't need to formulate the worry in terms of a dilemma since the second horn (that natural selection in fact tracked mind-independent moral truths) is so implausible. Consider Vavova's (2015) helpful discussion of evolutionary debunking arguments. Vavova correctly points out that to be a genuine evolutionary debunking argument, an argument must meet three conditions: (1) it must be *empirical* – that is, it must require some empirical claim about how evolution influences our moral judgments, lest it be a form of general skeptical argument, (2) it must be *targeted* – a debunking argument against moral realism should not also debunk our non-moral beliefs, and (3) it must be *epistemological* – a debunking argument does not claim that there are no moral truths, but that if there are, we do not or cannot know them (Vavova 2015, 105). Thus, vitally, a debunking argument is not a *general* skeptical argument. General skeptical arguments give reasons to *not be absolutely certain* of some domain of inquiry, but provide no positive reasons to actively doubt our currently-held beliefs in that domain of inquiry. For example, Descartes' evil demon might influence us to believe that we cannot be absolutely certain that the world is roughly as we experience it, but it does next to nothing in the way of providing evidence that we *should* actually believe that there is an evil demon manipulating us. Debunking arguments, Vavova rightly claims, aim to do more than cast a mere shadow of doubt: they aim to provide justified reasons to think that error is highly likely.²⁷

²⁷ Thus, debunking arguments are more similar to arguments like Bostrom's "simulation argument" than to Descartes' Evil Demon: where the Evil Demon argument aims merely to show that we cannot be absolutely certain that radical skepticism is false,

3.2.2 *The Radical Debunking Argument*

We are now in a position to develop a new, more radical version of evolutionary debunking arguments:

1. **(Possibility)** At the beginning of moral inquiry, the space of epistemically possible objective moral facts is exceedingly vast.
2. **(Evolutionary Influence)** Our moral minds are influenced, guided, shaped, and constrained by our evolutionary history.
3. **(Empirical Principle)** The mere fact that a (putatively) concrete object, property, or relation is referenced in a conceptual truth does not imply that the named object, property or relation exists concretely in our world.
4. **(Epistemic Hook)** To make progress past the beginning of objective moral inquiry, we need some epistemic hook into objective moral reality. Three main options for epistemic hooks are (a) fundamental moral conceptual truths, (b) a special moral epistemic capacity, or (c) epistemic tools and technologies to help narrow in on objective moral reality. However
 - a. **(Conceptual Truth)** Given (3) the Empirical Principle, we should doubt that fundamental moral conceptual truths can serve as an epistemic hook.
 - b. **(Capacity)** Given (2) Evolutionary Influence, we should doubt that any capacity of moral perception we have would accurately track objective moral reality
 - c. **(Narrowing)** Given the previous arguments (4a and 4b), we should doubt that the epistemic tools and technologies we use in other domains of knowledge could be helpful in narrowing in on objective moral reality.
5. Therefore, if there are objective moral facts, we should doubt that we will ever discover them. In other words, we should believe in moral anti-realism.

We'll need to take a closer look at this argument. To follow the argument, it is helpful think of it as contrasting moral inquiry from other types of epistemic inquiry that actually can succeed. The idea is that inquiry into any domain starts with an exceedingly vast number of epistemically possible facts (facts that, for all we know, could be true), but we use a variety of capacities, and intellectual and material tools and technologies to help narrow down this vast number of epistemically possible facts into something more manageable, and when successful, eventually acquire knowledge. At

Bostrom's simulation argument aims to provide reasons that we should believe radical skepticism.

the beginning of inquiry into meteorology, for example, it was a live epistemic possibility that gods were causing thunder. As inquiry progressed, this epistemic possibility was eliminated. The argument here goes that we have strong reasons to doubt that the capacities, tools, and technologies that we have used in our successful inquiries in non-moral domains will also work in the moral domain: we can never successfully move past the very beginning stage of (objective) moral inquiry.

Now we're in a position to unpack the argument and look at the premises in more detail. The first premise, **Possibility**, states that at the beginning of moral inquiry, the space of epistemically possible objective moral facts is exceedingly vast. A claim is epistemically possible if it is possibly true given a certain base set of knowledge (i.e. it is consistent with that base set of knowledge): it could be true given what we already know (whether we are likely to ever consider it or not). Claims that are not metaphysically or nomologically possible have been, for certain people at certain times, epistemically possible (it is not nomologically possible that thunder is created by the gods, or that phlogiston is a key element in combustion, but for certain people at certain times, these propositions were epistemically possible given those people's knowledge base). Even claims that are not logically possible, such as the law of non-contradiction, have been, for certain people at certain times, epistemically possible. In the beginning of inquiry in any domain, the number of epistemically possible facts *in that domain* is exceedingly vast (likely uncountably infinitely vast).²⁸ There is no reason to think that

²⁸ Since there are uncountably infinite mathematical facts, and the vast majority of domains of inquiry we are interested in are consistent with these mathematical facts, it is trivially true that at the beginning of inquiry into nearly any domain we are interested in, there are uncountable infinite epistemically possible facts. The claim here is stronger:

the moral domain is different in this regard. We need to appreciate just how vast and deep the field of epistemically possible objective moral truths is at the beginning of inquiry into the moral domain. First, and most obviously, at the very beginning of moral inquiry, any of the major normative ethical theories are epistemically possibly true: it is epistemically possible that we ought to maximize pleasure and minimize pain for the maximum number of beings who can experience pleasure or pain, or that we ought to develop the capacities and dispositions to most excellently fulfill our conspecific functions or that we ought to obey the commands of our god or dutifully follow the demands of the Kantian categorical imperative. And of course, normative theories that are unpopular, such as radical egoism, or even those that seem bizarre, such as a normative theory that asserted the only valuable activity in the universe was stamp-collecting are genuine epistemic possibilities at the very beginning of moral inquiry. But, the space of epistemic possibilities is much vaster: at the beginning of moral inquiry, even basic moral claims that seem true on entirely conceptual grounds such as “unnecessary suffering is bad,” “goodness and rightness are somehow related,” or “goodness and badness are not identical” are epistemically possible. We might be able to show that these claims are no longer epistemically possible almost immediately in the process of moral inquiry, but at the very beginning of moral inquiry, before inquiry has commenced, they are all epistemically possible. But, I must emphasize, the space of epistemic possibilities at the beginning of moral inquiry is *still* more vast than this: at the beginning of moral inquiry, it is epistemically possible that moral facts can be inconsistent with each other or that moral facts are ineffable or unknowable. Again,

that, for most domains, at the beginning of inquiry, there are uncountable infinite epistemically possible truths within the domain of inquiry.

this is not a special feature of inquiry into the moral domain – inquiry into just about any domain will share this feature. At the absolute beginning of inquiry into the domains we now think of as chemistry, biology, or meteorology, it was epistemically possible that facts vastly different from those we now believe are true were true, or even that inconsistent facts within each domain were simultaneously true. The key point is that at the beginning of moral inquiry, like at the beginning of other forms of inquiry, the space of epistemically possible truths is incredibly vast.

The second premise in the radical debunking argument, **Evolutionary Influence**, contains nothing that we have not already seen: that our moral minds are shaped by our evolutionary history, that our perspective of the space of epistemic moral possibilities is influenced and constrained for us by the way we evolved, and that, most importantly for us, only an infinitesimal portion of the epistemically possible moral facts at the beginning of moral inquiry will ever be live epistemic possibilities for us, given the kinds of creatures we evolved to be. Since this claim has been defended extensively elsewhere by supporters of traditional debunking arguments (Street 2006; Joyce 2007; Vavova 2015), I will not say much more about it than I already have here. It should be emphasized, however, that this claim does not entail that our moral beliefs are *only* those that were most adaptive in our relevant ancestral environments, that *all* human beings have shared extremely similar moral beliefs because of their shared evolutionary ancestry, or that we *only* ever act on evolutionary impulses. Evolution and biology are messy and produce tendencies more so than necessities. Our moral minds are not set in stone, and there is and has been a great variety of moral belief and attitude systems. Yet, it would be a mistake to infer from this that all or even a sizable portion of

the epistemically possible moral facts have been live possibilities for us – though the size of actually-held moral belief and attitude systems is vast, it absolutely pales in comparison to the size of epistemically possible moral facts at the beginning of moral inquiry. For example, some epistemically possible facts at the beginning of moral inquiry might be too complex to be live possibilities for us – imagine a 20,357-valenced system of scalar moral value instead of our two- or three-valenced one (morally good, neutral, or bad). Some epistemically possible moral facts might make claims we just wouldn't imagine given the way we evolved to interact with the world – perhaps that using our sense of the earth's magnetic field to navigate is morally wrong (which, lacking a sense of the earth's magnetic field, we wouldn't naturally come to imagine). And of course, some epistemically possible facts at the beginning of moral inquiry make claims that, given the attitudes we evolved to have, would be extremely counterintuitive or even repulsive to us.²⁹

The third premise in the radical debunking argument states the **Empirical Principle**: that the mere fact that a (putatively) concrete object, property, or relation is referenced in a conceptual truth does not imply that the named object, property or relation exists concretely in our world. It is an empirical question whether some object, property, or relation exists concretely in our world. The Empirical Principle is justified independently of debunking arguments. I take it to be a general ontological or epistemic principle. In its ontological form, which is the form used in the radical debunking

²⁹ This discussion is somewhat similar to the move Street (2006) makes in focusing on our “basic evaluative tendencies” instead of our entire moral psychologies. I do have a worry about this type of limitation: evolution directly or indirectly influences more aspects of our moral psychologies and our moral deliberation than our basic evaluative tendencies alone.

argument, it states simply that the fact that a (putatively) concrete object, property, or relation is named in a conceptual truth does not alone imply that the named object, property or relation exists in our world. In the epistemic form (which follows fairly straightforwardly from the ontological form), our confidence that a putatively concrete object, property, or relation referenced in a conceptual truth exists in our world should depend on our confidence, independent of that conceptual truth, that the objects, properties, and relations referenced in that conceptual truth exist in our world.³⁰

It will be helpful here to look at a few examples to help justify and understand the Empirical Principle. First, consider the infinite number of formal mathematical and logical systems that contain theorems and tautologies but appear to have no bearing on our world (for example, systems with vastly different rules of inference or arithmetical results than classical systems of logic or arithmetic). We can construct such systems easily: system X contains the digits 0-9, +, and =. It has only one axiom: for any digits a and b , $a + b = ab$ (that is, a concatenated with b). So, in system X, ‘ $1 + 3 = 13$ ’ is a conceptual truth (it is an instantiation of an axiom) – but this conceptual truth seems to lack much epistemic usefulness in our world: unlike our standard system of arithmetic, it does not help us solve problems we care about. The point is likely easier to see with fictional objects interpreted non-fictionally. Consider the following two conceptual truths:

Hobbit: Hobbits from Bree and hobbits from the Shire are hobbits.

³⁰ A slightly weaker version states that our confidence that a conceptual truth is epistemically useful for our world should depend on our confidence that the objects, properties and relations referenced in the epistemic truth closely *resemble* objects and relations in our world. Thus, geometric objects like perfect triangles or circles might not exist in our world, but insofar as objects in our world resemble them, conceptual truths about geometric shapes are epistemically useful.

Unicorn: A unicorn has one horn.

Now, if we interpret the fictional entities non-fictionally, the sentences look like this:

Hobbit*: Hobbits (which exist concretely in our world) from Bree (which is a location in our world) and hobbits (which exist concretely in our world) from the Shire (which is a location in our world) are hobbits (which exist concretely in our world).

Unicorn*: A unicorn (which exists concretely in our world) has one horn.

Hobbit* and Unicorn* are clearly false. Specifying the sentences like this helps us see that the mere fact that some object, property, or relation is named in a conceptual truth cannot imply that the object, property or relation exists concretely in our world.

One might be tempted to object to the Empirical Principle with a famous philosophical counterexample - Descartes' *cogito* argument: I am doubting that I exist, but if I am doubting, I am thinking, and if I am thinking, I am existing, so I exist. The "I" in the conclusion of the argument is meant to be a concrete object that exists in our world – that is the whole point of the argument. Notice first that the argument is largely, but not entirely comprised of a series of conceptual truths: that conceptually, doubting entails thinking and thinking entails existing. This helps us see how the *cogito* is not a counter-example to the Empirical Principle. The chain of conceptual truths alone does not establish that I exist: the chain of conceptual truths, plus the empirical information that doubting, thinking, and existing occur in our world, plus the empirical information that "I am doubting" establish it. When the conceptual and empirical truths are combined, the empirical premise that "I exist" in our world is entailed and the argument goes through. On the other hand, consider certain ontological arguments for the existence of God. Conceptual truths about God, such as God is (divinely) perfect or that (divine) perfection requires existence or would be better to exist than not exist reference

entities and properties (God, divine perfection) whose existence they are trying to establish. Unlike the *cogito* argument, we lack independent empirical evidence of the properties or objects named in the argument other than existence in general. Without independent reason to believe in the concrete existence of the objects, properties, and relation that together purport to entail the existence of God, we lack reason to believe that the argument succeeds in establishing the concrete existence of God. Thus, while the *cogito* argument does not run afoul of the Empirical principle, ontological arguments do.

The fourth premise, the **Epistemic Hook**, claims that to make progress past the beginning of inquiry into objective moral reality, we would need some epistemic hook into it. Three major options for epistemic hooks would be (a) fundamental moral conceptual truths, (b) a special moral epistemic capacity, or (c) epistemic tools and technologies (intellectual or material) to help narrow in on objective moral reality. There is nothing special about these epistemic hooks – they have analogues in other domains of inquiry. Inquiry into logic and mathematics advance as we come to better understand certain conceptual truths. Inquiry into the physical spaces around us (e.g. the location of this or that river or tree or mountain) can advance because we have capacities for perception and memory. Inquiry into astrophysics can advance because we invented intellectual tools and technologies like calculus, and inquiry into microbiology can advance because we invented material tools and technologies like microscopes. If objective moral truths are anything like objective truths in other

domains of inquiry, moral analogues of these epistemic hooks would help us push past the very beginning stages of moral inquiry.³¹

But, the radical debunking argument claims, we should strongly doubt that any of these epistemic hooks actually hook into the domain of objective moral facts. Let's look at them one by one. First, we might think, along with ethical intuitionists such as Sidgwick (1981), Moore (1903), Dancy (2006), Shafer-Landau (2005), Huemer (2005), and Parfit (2013) that there are some fundamental moral facts that are self-evident, such as "murder is wrong" – it is built into the very meaning of *murder* that it is wrong. But, if we want to spell out this claim more fully (though not completely fully) in an objective sense, this amounts to the claim "Murder, which is a type of act that exists concretely in our world, is wrong, which is an objective property that exists concretely in our world." But, we see an immediate conflict here with the Empirical Principle. What independent evidence do we have, beyond the fact that moral properties like goodness or wrongness are referred to in these conceptual truths, that these moral properties exist concretely in our world? Put simply, we have *no* evidence that moral properties exist concretely in our world beyond our confidence that they do. Now, if not for various skeptical worries, we might not have had any reason to doubt that

³¹ For the argument to be decisive, it would need to go something this like: (1) There are multiple ways to progress in a domain of inquiry (2) None of these ways apply to inquiry into objective moral facts, (3) therefore we could not progress in inquiry into objective moral facts. But, obviously we could not examine *all* possible ways to progress in a domain of inquiry, so we could not support premise 2 of this argument. Instead, I rely on a weaker, but still substantial claim: that the most likely (given their success in other domains of inquiry and their support among realist moral philosophers) options cannot work. Barring some other plausible option being suggested, we should doubt that we have any epistemic hook into objective moral reality.

confidence – but given the skeptical arguments we have been discussing, it is highly doubtful that this confidence is warranted.³²

The second possible epistemic hook is that we have a special faculty that tracks objective moral truth. Many philosophers (including Shaftesbury, Hutcheson, Hume, Smith, and Reid and the Confucians Mengzi and Wang Yangming) have claimed that we have special faculties of moral perception (though, importantly, not all of these philosophers believed that that sense was tracking mind-independent moral reality). But, this possibility, that we have faculties of moral perception that track objective moral reality is exactly what traditional debunking arguments call into question. Our basic evaluative tendencies, as Street (2006) calls them, seem importantly different from some of our other perceptive faculties. Many of our perceptive faculties, such as the ability to visually recognize objects or determine the direction of the source of a sound with our hearing, we might think, are accuracy-dependent. That is, the reason they developed the way they did was because they accurately detected things in the environment well enough to be selected for. This accuracy need not be perfect, and may

³² One might object against this argument by appealing to the earlier discussion of the *cogito*. After all, in that argument, which I said was successful, we start with an empirical fact about something that exists concretely in our world (my doubting), tie it conceptually to thinking and then to existing, and infer my concrete existence in this world. Why doesn't the moral version of the argument go through? After all, murder exists concretely in our world, and if murder implies wrongness just as doubting implies thinking, it seems the moral version of the argument must go through. But, there is a mistake here. To see it, imagine the argument *cogito**, according to which we have defined doubting as a mental activity of a physical body and thinking as the mental activity of an immaterial soul. So defined, *cogito** runs afoul of the Empirical Principle. We have concrete evidence of a physical brain thinking but not of an immaterial soul, so if we followed *cogito** we actually should not infer that we are thinking because we are doubting. Similarly, in the moral case, the intuitionist adds objective, mind-independent moral properties into the mix without providing any independent evidence that they exist concretely in our world.

actually be especially prone to false positives in the cases of risks, but detecting things that are really there is an ineliminable part of the explanation for why we have these capacities. But, in the case of our moral minds, we cannot so easily make the same claim. It seems highly likely that we would have evolved the same basic evaluative tendencies we have even if they tracked objective moral reality extremely inaccurately (Joyce 2007). There simply is no evidence that our basic evaluative tendencies are accuracy-dependent on objective moral reality.³³

Now, we need to make a (charitable) point that has been somewhat neglected in the literature on the debunking argument: that focusing so much on whether we have the *capacity* to acquire knowledge of objective moral reality is slightly misguided. After all, in many domains of inquiry, we have the (intellectual) capacity to discover pieces of knowledge that we never actually do discover. And in many domains of inquiry, we acquire knowledge not merely by means of our natural capacities, but because we have developed (intellectual and material) epistemic tools and technologies that, when combined with our natural capacities, help us to acquire new knowledge. This point opens up new possibilities for those seeking to defend moral realism against evolutionary debunking arguments: even if we did *not* evolve a faculty or capacity that directly tracks features of objective moral reality, that does not mean that we never could track them.³⁴ After all, we did not evolve specific capacities to directly see

³³ This is not to say that they are not accuracy-dependent on anything – they *do* track things like pain and suffering in others, acts of cheating and unfairness, and affronts to status or respect because of their role in fitness in our relevant ancestral environments.

³⁴ It is important to see how this line of argument is distinct from the response to debunking arguments that we saw earlier: that we evolved a capacity for rationality, and rationality allows us to discover moral truths. This line of argument is based not on our natural capacities but on technologies we invent: it claims that moral truths might be

microbes or solve differential equations, but we can do either utilizing epistemic tools and technologies that we invent. Perhaps we could discover objective moral facts the same way. This possibility, as far as I know, has not been explored as far as it should be in discussions of and responses to traditional debunking arguments. But, I think, this line of response does not seem ultimately viable to me, especially given what we've seen about other potential epistemic hooks. The successful epistemic tools and technologies we invent work because they improve upon, extend, and depend upon cognitive and perceptual capacities we already have: but if we have no epistemic hooks into objective moral reality, there is simply nothing to improve or extend upon. If our epistemic tools and capacities depend upon fundamentally veridically misinformed and misguided capacities, they will simply extend the misinformation.

3.2.3 Responses to Debunking Arguments

Having taken a closer look at the radical debunking argument, we are now in a better position to see how the main strands of replies to debunking arguments fail. There are at least four main strands of reply to debunking arguments: partner-in-crime accounts, confidence replies, third-factor replies, and narrowing replies. According to partner-in-crime accounts, evolutionary debunking arguments fail to meet Vavova's "targeted" condition. Thus, debunking arguments are too strong, and if true, would prove more than they should. For example, Justin Clarke-Doane (2012) has argued that if evolutionary debunking arguments of moral realism go through, then they should also similarly undermine confidence in objective mathematical truths. Our capacities to discover mathematical truths were also, like all our capacities, deeply influenced by

undiscoverable by our natural capacities (including rationality) alone, but discoverable with the use of the correct technologies.

evolution, so if our natural capacities undermine our ability to discover moral truth, so too do they undermine our capacity to discover mathematical truth. But, we are confident that we have discovered mathematical truth, so by modus tollens our evolutionarily-influenced epistemic capacities cannot undermine our ability to discover objective moral truths.³⁵ Now, we can see easily how partner-in-crime replies run into immediate trouble when aimed at the radical debunking argument, because the radical debunking argument provides a ready-made explanation of how inquiry in the moral domain is different from inquiry into nonmoral domains. In the case of mathematics, it is not the Evolutionary Influence premise but the Empirical Principle premise that is relevant. It was never in question that we could create systems that generate logical truths – the key question is which of those conceptual truths are describing anything that actually exists or happens in our world. That is an empirical question, and in the case of mathematics we have enormous amounts of empirical evidence that the systems we use *do* describe or model events in our world: because they work. There are other systems that do not work when attempting to describe or model events in our world. However, moral knowledge cannot be so utilized (moral knowledge is not even in the business of description and modeling) and so the radical debunking argument does not problematically undermine non-moral areas of knowledge.

According to confidence replies, we should not trust the conclusions of evolutionary debunking arguments because we are more certain of the truth of moral realism than of any of the premises in debunking arguments (see for example Nagel 2012, 105). Vavova (2015) points out one major issue for confidence replies: they serve

³⁵ That is, if moral realism is debunked, then mathematical realism is debunked. But mathematical realism is not debunked, therefore moral realism is not debunked.

best as responses to general skeptical arguments, and debunking arguments (whether in traditional or radical form) are not general skeptical arguments. As Vavova rightly points out, debunking arguments are closer in kind to an optometrist telling you, after a series of tests, that you are red-green color blind than to a stranger telling you that you see all colors incorrectly (2015, 105). The stranger, like the general skeptic, at best gives you a reason to be less than absolutely certain in some domain of knowledge, while the optometrist, like the debunker does much more than that: she gives you positive reasons to doubt what you took to be knowledge. The mere fact that I am highly confident in the truth of moral realism, perhaps more confident in it than any of the premises in the debunking argument cannot be used as a defense when that confidence and our warrant in it is being legitimately called into question.

There is another issue here with confidence replies. Once we specify what they claim, it is not at all clear that we actually are confident specifically in the truth of objective moral realism. I will happily grant that most of us are, when not thinking philosophically at least, incredibly confident in certain moral claims. But, are we confident that some moral claim is true, or that it is *objectively* true – that is that it is true completely independent of our evaluative attitudes and the type of creatures we are? A comparison will be helpful here. Consider “name realism,” an ontological theory (drawn from works of speculative fiction and fantasy such as (Rothfuss 2008)) according to which every type of object in the universe has an objective name that is part of the mind-independent fabric of the world. When someone argues against name realism, saying that it provides for an awfully ontologically bloated universe full of seemingly epiphenomenal names that seem perfectly explainable in terms of ordinary

social constructs, and one responds “Well, I am much more certain that this is called a “chair” and that is called a “table” than I am certain of any of your objections to name realism, and so I must accept name realism,” we should think that the name realist isn’t making a very good argument. Similarly, we should also think the confidence response doesn’t provide a very good argument, before we even start to worry whether it only succeeds as a response to general skeptical arguments.

According to third-factor responses, our moral faculties did not evolve to directly track moral truth, but some other faculty (a third-factor) *was* selected for, and that third factor *does* track moral truth. For example, a Kantian might reply that our discovery of moral truths depends on our capacities for abstract rational reflection, and these capacities were evolutionarily advantageous enough (their development helped us solve various practical problems that arose) to have been selected for. Radical debunking arguments problematize third-factor accounts along two lines. First, we need independent evidence (and such evidence is not forthcoming) that the third factors themselves are not epistemically distorted by our evolutionary history.³⁶ Second, even if they are *not* distorting themselves, if combined with the rest of our moral minds, there is little reason to think they could lead us to objective moral truths (imagine a person who followed perfectly truth-preserving rules of inference in her reasoning, but whose assumed premises were all false). Now, it is worth tackling third-factor responses where the third factor is reason or rationality in slightly more detail here. I do not deny that we evolved a faculty to use reason, to follow systems of inference, and that in many epistemic domains, this ability is *ceteris paribus* a reliable belief-forming mechanism

³⁶ This will be especially important for third-factor replies in which the third factor is essentially moral (e.g. a concern for justice).

(or more accurately, a mechanism that contributes to reliable belief formation). We have evidence from non-moral domains of inquiry that some potential third factors, like rationality, are not epistemically distorting: rationality, when applied carefully, works perfectly well in mathematics. Thus, the first response I gave to third-factor accounts does not apply to them. However, the second response remains. We are not justified in inferring that rationality alone is also a reliable belief-forming mechanism in the moral domain. First, suppose that moral truths must follow certain reasonable principles of rationality: they must be consistent and some set of rules of inference similar or identical to the rules of classical logic applies to them. But this could only guarantee that we could eventually arrive at a consistent set of moral beliefs and a consistent set of beliefs is not necessarily a true set of beliefs. Correct outputs of reasoning rely on correct inputs, and the debunking argument provides strong reasons to doubt that the values we begin with are *objectively* correct moral values at all. Let's consider an example.³⁷ In our adaptive history, developing a taste for sugar was highly adaptive: it was a good source of energy that was worth seeking out, and rare enough that overconsumption would not lead to fitness-decreasing problems. But today, sugary food and drinks are widely available. However, using reason, we recognize that our evolved basic tendencies towards sugar are problematic, and we override them. And so, we come to have a response to debunking arguments. But, this line of reasoning does not show that reason has helped us discover an *objective* moral truth. We have not discovered that sugar is *objectively* bad for us, or even *objectively* unhealthy in the strong sense of "objectivity" we are interested in here. All we can say about sugar is

³⁷ Thanks to Nancy Snow for suggesting discussing this example.

that the consumption of certain amounts of it *ceteris paribus* tends to have certain bodily effects on humans, such as weight gain or tooth decay. It requires another step to say that weight gain or tooth decay are objectively bad, though they very well might impair some functioning. But then we need an argument to do all this, and I just do not see how we can construct one without surreptitiously relying on our practical perspectives and the attitudes within them, and our practical perspectives, while not determined wholly by natural selection are always shaped and constrained by it.

Even if we grant, despite its exceedingly small probability, that some of the values we begin with are objectively correct, we should also grant (given their internal inconsistency) that some of them are not. Following reason, in this case, will not guarantee that we will eliminate the objectively incorrect values – we might arrive at consistent moral systems by rejecting some of the inconsistent objectively correct moral values. But, the problem is much deeper than all of this: if reason amounts to the ability to follow certain patterns of inference, how could we identify the correct patterns of inference at the beginning of inquiry into the objective moral domain? This is the problem we have run into with the Empirical principle. We might assume that moral truths are consistent with each other, but even this most basic assumption has no objective evidence at the beginning of moral inquiry. Further, note an asymmetry here between moral inquiry and other types of inquiry. In other types of inquiry, we quickly uncover justification for requiring consistency: as we understand the world better, and develop fuller explanations of the way things are, it becomes apparent that the best explanation for the successes we achieve in assuming that truths must be consistent is that in nature, there are no contradictions. But, we need evidence that morality is like

this. It might not be, and we lack evidence that it is. If we cannot determine what system of inference to apply to moral claims, it is unclear how reason alone could help us form reliable objective moral beliefs.

Narrowing responses reject the Possibility premise in debunking arguments, which I treat in the radical debunking argument as the claim that “at the beginning of moral inquiry, the space of epistemically possible objective moral truths is exceedingly vast.” They claim that the scope of epistemically possible moral truths is actually *not* exceedingly vast, because we have access to several conceptual moral truths, such as “murder is wrong” or “it is good to do the right thing.” But, given the Empirical Principle, this line of argument faces severe problems. Even if we grant that it is an analytic, conceptual truth that “murder is wrong” (which, I am quite happy to grant) it does not follow that fully objective wrongness (and thus, murder, so-defined) exist concretely in our world. But for moral realism to be true, they must. By way of comparison, narrowing arguments look an awful lot like a group of misguided scientists claiming that they are making genuine progress towards mapping the genome of a unicorn because they have narrowed down the scope of their search for unicorns (the first stage of their project) by eliminating all animals that did not resemble white horses with one horn on their heads.

We should now take a step back and appreciate some of the implications here. First, there is an especially major challenge for intuitionist versions of moral realism. We have reasons to doubt even that the most basic putatively conceptual moral truths are *objectively* true. Second, versions of moral realism based on special faculties of moral perception do not fare much better. So, the radical debunking argument might

suggest that the best hope for moral realism is that moral truths are like calculus or Gödel's incompleteness theorems or nuclear fusion, discoverable using created epistemic tools and technologies. But, it is not at all clear what epistemic technologies could lead us to mind-independent moral truth. And so, even if moral properties and entities were woven into the very fabric of the universe, the probability of finding them is infinitesimal. On the other hand, treating morality as something we invent, and construct, and reinvent, and revise gives us something to do. Seeking the objective truths that moral realism insists are out there, somewhere, leaves us lost and adrift, rudderless and directionless in a strange and alien moral cosmos, unfathomed and perhaps unfathomable, with no hope of ever finding our way.

3.3 Constructing Morality

3.3.1 Metaethical Constructivism

In the last section, I presented a radical debunking argument against moral realism and argued that if moral realism were true, we could not ever discover objective, mind-independent moral truths. But, if moral realism is false, should we instead assume that some strong version of moral anti-realism or nihilism is true, according to which *nothing* is valuable and all moral claims are always and essentially false? Doing so would be far too hasty, and in fact drastically out of line with the arguments we just saw. After all, a key claim in the radical debunking argument is that we already are valuing and devaluing creatures, that our moral psychologies are in large part, but not totally, given to us, and that we do not discover morality out in the world but put it there ourselves. If this is the case, a perspective shift is required. In claiming that there is no objective value in the world and no objective moral truths,

nihilism aims to pull a rug out from underneath us that we were never really standing on. Moral claims, we should think, do not and never did acquire justification or normative force from features of the world entirely outside our minds. We create moral systems from our values for our circumstances and constantly refine them to better or more effectively express our values while at the same time reflecting on and refining our values themselves.

If something like this view is correct (and we'll spell the view out in more detail shortly), we should briefly note two points about how it problematizes a much wider range of metaethical theories than simple and strong versions of moral realism and anti-realism. First, because morality is a system we construct and constantly reconstruct and refine, we should not be very interested in metaethical views that purport to provide an analysis of some fixed meaning of moral terms and claims, such as realist cognitivist views according to which our moral claims report our truth-apt moral beliefs (e.g. Shafer-Landau 2005; Huemer 2005), or anti-realist non-cognitivist / expressivist views according to which moral claims express our non-truth-apt emotional reactions (Ayer 1952) or norms or attitudes (Gibbard 1990) - the meaning of moral claims depends on the ways we have constructed and understand our systems of morality, and so the meaning of moral claims at any particular time is contingent and changeable. Similarly, fictionalist metaethical views (e.g. Mackie 1991; Joyce 2007) according to which moral terms like "good" or "right" mean *objectively* good or right, and thus falsify any moral claim containing them (because objective goodness or rightness does not exist) are problematized: if people would use moral terms understood as constructions rather than as referring to objective features of the world, those terms would not falsify the

sentences containing them. The second point we should briefly note, which we will also discuss in more detail shortly, is that the perspective shift away from concern over the existence of objective moral truths is also a shift away from a narrow concern about moral propositional attitudes. We do not just bring our moral beliefs to the table - we bring our entire moral psychologies, including patterns of perception, inference, salience, interpretation, association and explanation. Thus, a moral epistemology that focuses only on our beliefs (and their truth and justification) should seem oversimplified and outdated: our moral minds are far richer and far messier, and our metaethics and moral epistemology should reflect our best understanding of that richness and that messiness.

But, fortunately, we do not need to start from scratch here, in at least three important ways. First, when I claim that morality is a system we invent and constantly refine based on our values, this is *both* a normative and a descriptive claim. It is normative in the sense that it tells us how we ought (epistemically *and* morally) to think about morality, and upon examination and reflection gives us directions for justifying and improving it. But, it is also a descriptive claim in that, if the epistemic premise that there are no knowable objective values and the psychological premise that we are valuing creatures nonetheless both hold, then constructing and reconstructing, designing and redesigning morality *is exactly what we have always already been doing*. We have developed the ethical norms, principles, institutions, theories, practices, and perspectives we have by refining our moral systems based on those antecedently-held values and perspectives. Thus, the account of moral justification that I will develop here

will and *should* look extremely similar to the ways that we already attempt to give moral justification.

Second, this is not totally new territory in moral philosophy. A wide variety of ethical theories have explicitly treated and provided robust accounts of morality as a system we have invented to better express and support our values and achieve our goals. These include, to take a small sampling, fictionalist projects such as those of Mackie (1991) and Joyce (2007), pragmatic naturalist theories such as those of Kitcher (2011, 2012) and Sarkissian, Wong, and Flanagan (2016) and the “accommodationist” anti-realist views that developed in dialogue (Campbell and Kumar 2012, 2013; Kurth 2017), and Humean constructivism as developed by Street (2010, 2006; see also Arruda 2017). They also include historical constructivist moral theories, the earliest of which that were developed in detail likely were those of the early Confucian philosophers Xunzi and possibly Mengzi.³⁸

Third, as we’ve seen, a key feature of the metaethical view I’ll develop here has been its commitment to a realistic moral psychology: it eschews simplified, reductive models of our moral mind comprised entirely of our moral beliefs and desires and instead treats our moral minds in terms of perspectives. This is, in some ways, a radical

³⁸ One major strand of metaethical naturalism, Aristotelian ethical naturalism, claims that there are objective natural facts about human functioning that ground objective normative facts (Hursthouse 1999; Foot 2001; M. Thompson 2008). On my own view, the normative facts grounded in claims about human functioning generated by these theories are not objective in the strong sense we have been using here, though they may be objective in a weaker sense. The main reason I think this, though I will not defend this view at length here, is that I am highly skeptical that any putatively naturalistic account of human functioning does not sneak in perspectival normative features, and more tendentially that there could be no naturalistically respectable account of human functioning as anything other than a vessel for our genes and that such an account cannot successfully ground the normative views that Aristotelian ethical naturalists support.

departure from simplistic models of our minds that have been historically common in ethics. But here we also do not need to start from scratch. There has been a growing literature examining perspective and the role it plays in our epistemology (see Camp 2006, 2013; Elgin 2002, 2010; Riggs 2016). But before we discuss that literature in more detail I will need to say about more about the positive metaethical view that I am defending in this dissertation.

3.3.2 Perspectival Naturalism

Let's pause to check our bearings here. So far, I have argued that strong versions of moral realism and anti-realism (as well as several other metaethical views) cannot succeed, and that morality is a system we create and refine over time based on values that were originally given to us largely but not entirely by evolution. The next step of the project is to describe in more detail the alternative metaethical view I will adopt, perspectival naturalism. As I said at the outset of this chapter, we need this metaethical view to do three things. The first is that it must be plausible - especially given the radical debunking argument and our best scientific understanding of our moral psychologies. Relatedly, the second thing that perspectival naturalism must do in this chapter is provide a framework for how moral claims, systems, and perspectives are justified because doing so is both a requirement for any plausible non-nihilist metaethical view and will be necessary for providing guidelines in how the normative argument in the next chapter for the Consequences and Character conditions should go. The third thing that perspectival naturalism must do in this chapter is that, without running afoul of our second goal, it must somehow dislodge our currently-held moral intuitions. It must provide reasons to accept the relatively revisionary nature of the

Consequences and Character conditions without reverting to some radical form of normative subjectivism.

Street (2010) helpfully distinguishes between two versions (Kantian and Humean) of metaethical constructivism. They share the idea that “the truth of a normative claim consists in the claim’s being entailed from within the practical point of view” (Street 2010, 367) and “the subject matter of ethics is the subject matter of what follows from within the *standpoint* of creatures who are already taking this, that, or the other thing to be valuable” (Street 2006, 367 emphasis added). So far, this is entirely compatible with the radical debunking argument. Kantian constructivism further claims that certain substantive normative truths are entailed from within *any* practical standpoint, whether we realize it or not. The view’s namesake gives a fine enough example to aid in understanding: Kant believed that justification within all moral standpoints must be rational and that features of rationality entail the categorical imperative (in its three equivalent forms). Kantian constructivism gives these universally-entailed moral claims a tint of objectivity, though importantly for Street applying the term “realist” to Kantian constructivism is not entirely accurate: normative truths still are grounded ultimately in features of moral minds and not independent of them (Street 2010, 371). Humean constructivism, on the other hand, claims that no substantive moral truths are entailed from all practical points of view (Street 2010, 370). Street suggests, correctly, that debunking arguments should point us towards this Humean constructivism (Street 2006, 2010).

There are several reasons that debunking arguments present challenges for Kantian constructivists, and we should briefly mention a few here. First, let’s suppose

for the sake of argument that some particular set of normative claims *is* entailed from within all moral perspectives. But, if our moral perspectives have been given to us largely by evolution, and evolution itself does not track moral truths, why should the fact that some normative claims are entailed from any moral perspective serve in justifying them? If the systems themselves are not essentially truth-tracking, the most likely explanation for the fact that some claim is entailed in all of them is not that it is somehow more epistemically justified or more likely to be true but instead that was incredibly adaptive (Street 2010). Relatedly, we might ask what *extra* justification the fact that some moral claim is entailed in any moral perspective provides over and above the fact that it is entailed from the moral perspective a person currently occupies. If we think normative justification can come from *within* moral perspectives, the fact that some claim is entailed from within a separate moral perspective that I do not occupy might be beside the point (even if that moral claim is also entailed in many other or even all separate moral perspectives).³⁹ But, the most important line of objection to Kantian versions of constructivism question the supposition we began with – that some particular normative claim or set of normative claims actually is entailed from within all moral standpoints. The idea that any substantive moral claim could be entailed from within *all* moral standpoints is, not to put too fine a point on it, scientifically naïve and epistemically unjustified (see Street 2010, 2006). We simply have not encountered

³⁹ To be more clear, the criticism here will not be that we can only be justified in believing moral claims that are entailed from the practical standpoint we currently occupy – this would mean we never had a reason to change our moral minds and accept different moral systems. Instead, the worry is that IF a claim is rationally entailed in my perspective, and I am justified in believing it from my standpoint, then the mere fact that it is entailed in all other practical standpoints provides no *additional* justification for the claim than I already had.

enough creatures with moral standpoints different from our own to believe with justified confidence that it could be true. And, we have strong reasons to think that it could not be, if our moral perspectives are influenced by our moral psychologies and our moral psychologies are influenced by our ancestors' specific evolutionary history. Any apparent convergence of moral perspectives, the objection goes, is easily explainable: we have focused on human moral perspectives and humans have a tightly shared evolutionary history and thus a lack of range in our moral psychologies.

So, it seems that Humean constructivism is in a better position, given evolutionary debunking arguments, than Kantian constructivism. But, this comes at a cost. It would be very nice if we could be justified in believing in Kantian constructivism. It would give us something clear and optimistic to say about, for example, moral disagreement. Even if we disagree about some ethical issue, there *is* likely a solution to the disagreement that anyone, coming to the disagreement from any moral perspective could eventually come to understand, because some substantive moral truths are entailed from within *any* moral perspective. Further, Kantian constructivism can potentially give us a nice answer to the question "Why should we think that our current moral systems are justified?" A Kantian constructivist might answer that moral systems, given proper rational reflection, eventually converge on those substantive moral truths entailed from any moral perspective, and that our currently-held moral systems (at least, the more enlightened ones), while not perfect, are close enough that we are generally justified in trusting them.

In short, what all this means for us is that we need to add two more desiderata to our metaethical theory. We've already seen that we need our metaethical theory to be

(a) philosophically and scientifically defensible, (b) provide a general framework for justifying moral claims, and (c) justify the project of questioning or dislodging our strongly-held moral intuitions (without necessarily abandoning them altogether). We now adjust this list by modifying the second of these three goals: in addition to needing an account of how moral claims or systems can be justified, that account should also (d) give us a path forward in cases of moral disagreement and (e) provide the account of justification without appealing to the claim that certain moral systems are more justified because we will universally converge on them (given adequate deliberation, reflection, and information). To do all of this, we will need to start filling out some of the details of Humean constructivism.

We will call this filled-out version of Humean constructivism “perspectival naturalism.” It derives from Humean constructivism a commitment to a naturalistic worldview,⁴⁰ to metaethical anti-realism, and to a generally constructivist account of moral justification. Perspectival naturalism supplements Humean constructivism by adding two key commitments to the view: a commitment to what I’ll call the “design paradigm” and a commitment to an explicitly perspectival treatment of our minds. Like

⁴⁰ “Naturalism” is admittedly not the clearest term in all of philosophy. The term itself was developed largely in a metaphysical and epistemological context to pick out a philosophical worldview that aligned itself with science in trying to understand the world and human beings and that posited nothing “supernatural” (Kim 2003; Papineau 2015). Perspectival naturalism’s commitment to naturalism should be interpreted primarily (though not entirely) as a commitment to descriptive or empirical adequacy: insofar as our metaethical or normative theorizing relies on descriptive or empirical claims, we should ensure that we are relying on only the best justified descriptive and empirical claims, and typically that will require that these claims be scientifically informed. I will leave this commitment to naturalism largely undefended. The main reason for doing so is that there is little if anything I could contribute here that would sway those who hold entrenched views about naturalism one way or the other.

many anti-realist and constructivist views, perspectival naturalism is driven by the goal of providing an account of which moral theories or systems we should or should not adopt – which are better or worse justified than others – given that we cannot know objective moral truths, but do not want to (and given the kinds of creatures we are, cannot) abandon morality altogether. As we have seen, many of these views, including perspectival naturalism will attempt to do so in reference to our moral values and perspectives and who we are, morally: these factors, and not objective moral truth, are what must ultimately justify moral theories and systems. However, the details of exactly how this justification is derived for perspectival naturalism will be slightly different from competitor anti-realist or constructivist metaethical views. In particular, for perspectival naturalism, the best moral theories or systems for us are those that are best *designed* for us: those that best bring into alignment our numerous, various, sometimes mutually-supporting, sometimes competing moral considerations. These considerations will include familiar moral considerations such as standard moral values of concern for the well-being of others and respect for their dignity. But, vitally, perspectival naturalism also includes *design* considerations: we can evaluate theories not only by how well they capture our familiar moral values but also by how well they work for us, given our needs, abilities, limitations, and circumstances.

3.3.3 *The Design Paradigm*

The design paradigm is a way of thinking about what we do and should do when we are morally engaged with the world and each other. Recently and historically, several philosophers have developed, to varying extents, normative accounts of morality

as a system we create and refine based on our values and goals. In contemporary metaethics, this view was developed most famously by Mackie:

Morality is not to be discovered but to be made: we have to decide what moral views to adopt, what moral stands to take. No doubt the conclusions we reach will reflect and reveal our sense of justice, our moral consciousness - that is, our moral consciousness as it is at the end of the discussion, not necessarily as it was at the beginning. But that is not the object of the exercise: the object is rather to decide what to do, what to support and what to condemn, what principles of conduct to accept and foster as guiding or controlling our own choices and perhaps those of other people as well (Mackie 1991, 106)

One absolutely vital idea that pervades views like Mackie's is that recognition of the metaethical fact that moral systems are created by us has implications for which normative theories we should adopt. Of course, different philosophers will determine different implications for normative theorizing based on how they construe the created system of morality. Mackie (1991, 107) noted (but did not endorse) one historically important strand of theories that did so by treating morality as a "device for counteracting limited sympathies." This view of morality as a device, or a technology, with a particular function that will help frame and justify particular normative theories will be important for us to discuss and motivate here in some detail: for perspectival naturalists, these views are very much on the right track, and very nearly correct.

Expositing on Warnock's version of this view, Mackie writes:

Warnock explains [the object / function of morality] in terms of certain general and persistent features of the human predicament, which is 'inherently such that things are liable to go very badly' - badly in the natural, non-moral sense that human wants, needs, and interests are likely to be frustrated in large measure. Among the factors which contribute to make things go badly in the natural course of events are various limitations - limited resources, limited information, limited intelligence, limited rationality, but above all limited sympathies. Men sometimes display active malevolence to one another, but even apart from that they are almost always concerned more with their selfish ends than with helping one another. The function of morality is primarily to counteract this limitation of men's sympathies. We can decide what the content of morality must be by inquiring how this can best be done. (Mackie 1991, 107-8)

Of course, this isn't a particular new idea. Mackie points out this seems to be the idea behind the myth of morality told by Plato's Protagoras. According to that myth, Prometheus and Epimetheus were assigned the task of giving every creature its unique abilities (Protagoras 320d-321c). They accidentally used up all of the unique abilities before they got to humans, and so Prometheus stole from Hephaestus and Athena their technical (craft-making / building) skills and (more famously) the ability to use fire, and gave them to humans (Protagoras 321d). But, these gifts were not adequate:

Thus equipped, men lived at the beginning in scattered units, and there were no cities; so they began to be destroyed by the wild beasts, since they were altogether weaker. Their practical art was sufficient to provide food, but insufficient for fighting against the beasts-for they did not yet possess the art of running a city, of which the art of warfare is part-and so they sought to come together and save themselves by founding cities. Now when they came together, they treated each other with injustice, not possessing the art of running a city, so they scattered and began to be destroyed once again. So Zeus, fearing that our race would be wholly wiped out, sent Hermes bringing conscience and justice to mankind (Protagoras 322b-c).

Thus, according to this myth, the function of our moral sense or conscience (*aidos*) and law and justice (*dike*) was to reduce strife and conflict. Though Mackie does not discuss him, this parallels the Confucian philosopher Xunzi's account of the creation of ritual (a moral concept as central to Confucian ethics as justice was to the ethical systems of the Ancient Greeks) and rightness:

From what did ritual arise? I say: Humans are born having desires. When they have desires but do not get the objects of their desire, then they cannot but seek some means of satisfaction. If there is no measure or limit to their seeking, then they cannot help but struggle with each other. If they struggle with each other then there will be chaos, and if there is chaos then they will be impoverished. The former kings hated such chaos, and so they established rituals and *yi* [rightness] in order to divide things among people, to nurture their desires, and to satisfy their seeking. [Xunzi, 19.1-10]

Thus, for Xunzi, the justification for the moral prescriptions of rightness and Confucian ritual derives from their effectiveness in reducing strife and chaos and in nurturing people's wants to lead to satisfactory or satisfied lives. Mackie also places Hobbes and Hume in this camp (Mackie 1991, 108–11). For Hobbes (1998; Mackie 1991, 108–10), of course, the function of moral principles and the social contract is to ensure we can acquire safety and security, and the justification for the moral principles he suggests and the social contract derives from their putative ability to most effectively align our rationally self-interested incentives. And famously, for Hume, justice was an *artificial* virtue: “It is only from the selfishness and confined generosity of man, along with the scanty provision nature has made for his wants, that justice derives its origin” (Treatise III.II.II, qtd. on Mackie 1991, 110).

The most developed version of the view that the function of morality is to best manage our limited sympathies that I know of is Kitcher's (2011, 2012) “pragmatic naturalism.” Kitcher claims:

Ethics emerges as a human phenomenon, permanently unfinished. We, collectively, made it up, and have developed, refined, and distorted it, generation by generation. Ethics should be understood as a project – the ethical project – in which we have been engaged for most of our history as a species. (Kitcher 2011, 2)

Kitcher argues that the circumstances of our evolutionary history pushed us to develop a capacity for psychological altruism – that is, a capacity to help others out of a genuine concern for them – but that capacity is limited. Failures of psychological altruism occur, and moral systems develop and progress as they improve their effectiveness at increasing the benefits from psychological altruism and limiting the costs of altruism

failures. Thus for Kitcher, the correct way to understand the activity of ethics, both descriptively and normatively, is as a *technology* (Kitcher 2011, 7).

Perspectival naturalism shares (and is inspired by) more than the name only of Kitcher's pragmatic naturalism. Both theories share a commitment to a naturalistic worldview. Both treat morality as a system created by us and claim that recognition of this fact should inform our normative theorizing (though in different ways). Both claim that there is no single perfect, moral system – that ethics is “permanently unfinished” (Kitcher 2011, 2), though again for different reasons. Finally, both views take moral justification to work largely as we already treat it and have treated it– that is, the product of our antecedently-held moral values and well-informed ethical deliberation. We should, however, explore some of the differences between the two theories. The first difference derives from a concern (and echoes a concern raised by Mackie (1991) as well) that treating morality as a system whose function is to manage our limited sympathies (or for Kitcher, failures of psychological altruism) is simply too narrow of a focus. Concern and sympathy for others' well-being is an important, maybe even fundamental moral consideration, but it is simply not the only moral consideration for us. We are concerned with justice and with fairness, with respect and dignity, with a great host of personal character traits like honesty or compassion or fortitude or humility or integrity that are related with but do not reduce without loss to a concern for others well-being. And, this should be expected: our moral systems grow out of our moral psychologies, and our moral psychologies are a hodgepodge of basic evaluative

tendencies that evolved at different times, in different circumstances, and for different reasons.⁴¹

For this reason, perspectival naturalists prefer not to treat morality as a technology, though doing so is highly informative and nearly correct. Treating morality as a technology makes the most sense when treating morality as having a particular, fairly narrowly-defined function – as most pieces of technology do. But, morality does not have a single, narrowly-defined function. Instead of thinking of morality in terms of technology, perspectival naturalism treats morality in terms of *design* – that is, like software or user experience design, interior design, fashion design, layout, print, or graphic design, industrial, organization, or systems design, or building, spatial, landscape, architectural and service design. What these all have in common is that they aim at effectively achieving *various* goals and functions (some of which will be mutually supporting, some of which will be in tension), given our particular values, preferences, circumstances, perspectives, patterns of behavior and thought, resources, and capacities. Good design manages and aligns *all* of these factors as well as can be done. Thus, for a perspectival naturalist, one ethical theory is better than another not when it fulfills the (narrow) function of morality more effectively than another but

⁴¹ Spelled out in more detail, the criticism here is that these functional approaches to morality offer something of a bait and switch. These functional approaches purport to derive normative implications from the realization of the descriptive fact that all that morality is is a device with a particular function, and thus all that justification could be for some moral theory or other is how effectively they fulfill that function. This would be a very nice picture: it would give us our theories (the well-functioning ones, anyway) normative force but not at the risk of ontological bloat or a weird metaphysics. But, when we look closely, we see that these functional accounts, with their narrow treatment of morality as a device for counteracting limited sympathies or altruism failures *leave a great deal of human morality out*. But, this is actually a fatal flaw: the theories were premised on the claim that descriptively, morality *just is* a device with that function – but this is descriptively false.

when it is *better designed* than another. This is the design paradigm in perspectival naturalism.

But, what will this actually mean for doing ethics? When is an ethical theory *better designed* than another? We can take our cues here from actual examples of good design. We should begin with an example that helps illustrate how design *in general* should inform our ethical theorizing. Consider the case of the Frankfurt Kitchen, designed in the 1920s by Margarete Schütte-Lihotzky:



Figure 1: The Frankfurt Kitchen (view from the entrance). Public Domain.
https://en.wikipedia.org/wiki/Frankfurt_kitchen#/media/File:Frankfurterkueche.jpg

The Frankfurt Kitchen is the basis for the design of modern “built-in” kitchens and was even part of an exhibit in the MOMA: before the Frankfurt Kitchen, kitchens in Europe

and America were typically not professionally designed at all – they were just the location where lots of tasks (including dining, bathing, sleeping, and clothes-washing) were performed (“The Frankfurt Kitchen,” n.d.; Filler 2011). The design of the Frankfurt Kitchen did not have a single, narrow goal, but multiple goals of improving efficiency, workflow, hygiene, convenience, and well-being: it integrated the facilitation of these goals at the general level and in the tiniest details: it included flour containers made of oak to repel mealworms (most of the other containers were made of aluminum) and cutting surfaces made of beech to reduce accidental stains or knife marks; it included a window to let in sunlight to reduce the typical darkness (and thus, hopefully to slightly reduce the typical drudgery) of kitchens in traditional lower and middle class housing (“The Frankfurt Kitchen,” n.d.; Filler 2011). In meeting these goals, its design was informed by the best science in industrial design, by the assembly-line techniques of Ford, and by Schütte-Lihotzky’s own careful empirical research that included time-motion studies to test the efficiency of the design and numerous interviews to gather information about the preferences and behavioral patterns of users of the kitchen. The design was also shaped by various values: a drive for the modernization and rationalizing of home spaces, which was seen at the time as empowering and promoting egalitarianism (“The Frankfurt Kitchen,” n.d.). It was constrained in space and in material costs, and in how easily and cheaply its parts could be transported and assembled.

What the case of the Frankfurt Kitchen helps illustrate for ethical theorizing and the design paradigm is this: (and here, perspectival naturalism shows its constructivist roots) good design *begins* with the goals, desires, preferences, and needs of the users in

mind and aims to facilitate the acquisition of those. But, this facilitation is not the only factor that goes into good design: it is also *informed* and *constrained* by who we are: by our values and circumstances, our resources and capacities. A well-designed ethical theory, like a well-designed architectural space, brings *all* of these factors into alignment as best as possible.

Importantly, good design does not over- or underprioritize these various factors. Consider advances (if they can be called that) in gaming / gambling design and especially slot machines aimed at promoting compulsive or addictive use (“advances” which have begun to be utilized in other domains such as social media app and video game design). The central concept behind these designs is the “Skinner box” or “variable ratio enforcement,” stemming from an experiment in which Skinner found that rats spent more time pulling a lever when doing so variably gave out food pellets than when doing so always did: by varying the ratio of rewards and losses, and frequently providing small rewards, slot machine and social media app designers can keep their users using their products for longer and longer periods of time (Schüll 2014; Alter 2017; A. Thompson 2015). Other related design strategies have been developed: in slot machines, there are often multiple paths for rewards, including mini-games that are perceived as requiring skill to win, while the overall chance of winning something has increased to around 45%, reducing the feeling of risks (A. Thompson 2015). Now, it is tempting to say that these newer slot machines (and other technologies that use their design principles) are better designed than the products they have replaced. But this is too quick – they are only better designed given certain goals and values. Given other goals and values, they are much worse designed. Pursuing this intuition, these designs

overprioritize the goals of people hoping to profit from using psychological techniques to addict consumers to their product, and underprioritize the constraints from values against coercion, manipulation, and profiting from the harm of others. Manipulative design is not good design, we might think. Consider the concept of “Dark Design Patterns / Dark UX” – known more colloquially in development circles as “asshole design” – design practices (most famously used in, but hardly limited to smartphone / computer / web apps): for example, making a button that will cause you to subscribe / and or pay for a service *look like* a button that will cancel the service, or hiding an unsubscribe / cancel button entirely, or, in an entirely different domain misleading packaging such as food containers with shapes designed to hide that the container holds much less food than it initially appears to hold. Thus, just as designs of products can go bad by over-prioritizing and under-prioritizing certain goals and values, so too can some ethical theories be poorly designed.

There are many more specific ways in which examples of good (and bad) design can inform perspectival naturalism. We will not examine an exhaustive list of them here, but instead focus on those that will be especially important for the arguments I’ll develop in the next chapter. In particular, we will focus on three types of cases: (1) design that better utilizes our natural tendencies or antecedently-developed ways of doing things, (2) design that better directs our attention, (3) design that, in addition to facilitating the achievement of its intended goals, provides extra benefits or positive externalities. These cases are important not because they illustrate especially powerful moral considerations: in fact, we will see that the moral considerations they introduce are fairly weak, easily-defeasible, pro tanto moral considerations. But, what is important

about them is that they are moral considerations *at all*: while they are perfectly sensible and plausible moral considerations when we are thinking of morality in the design paradigm, they would not be (or at least, not necessarily be) otherwise.

First, consider the case of desire paths in urban landscaping design. Larry Wall, who invented the programming language PERL, said one of his goals in designing the language was to mimic something that had been done on the campus at UC Irvine:

People will accept a new thing much better if it already resembles something they're familiar with or the way they are already thinking about things. A musician would say "A musical piece lays under the fingers -- it looks hard but it is easy to play." Another way of thinking of it is [by analogy:] At the University of California at Irvine, when they first built its campus, they just planted grass. Then they waited a year and looked at where people had made paths in the grass and built the sidewalks there. (Dougherty 1998)

This story, at least as presented here, is (as far as I can tell) apocryphal – versions of it occur about several college campuses, including one version where Thomas Jefferson used this design strategy at the University of Virginia. But, the likely reason that the story is so ubiquitous is that it is actually a fairly common practice in landscape and urban design to sidewalk so-called desire-paths: pathways that humans and animals naturally take when moving through spaces. The takeaway for moral theory here is that a potential feature of a well-designed moral theory is that it will (when possible) work *with* our natural tendencies and antecedently-developed ways of doing things, rather than *against* them. Now, we need to be careful here. This claim could easily be misinterpreted as a kind of radical naturalism: that what we ought to do is simply what is most natural for us. This is not the claim at all. Many times, our values or goals will require us to act against our natural or antecedently-developed inclinations and tendencies, to work on and reshape them. The claim here is much weaker than that

radical naturalist claim. Instead, the idea is something like this: when two competing moral theories are equally or relatively equally well-designed in other areas (e.g. how well they capture our values or facilitate achieving our goals), if one theory better utilizes our natural and antecedently-developed tendencies and inclinations, then it is preferable.⁴² This is not an altogether new form of argument in ethics. For example, Joshua Green and colleagues (J. D. Greene et al. 2008, 2004; J. Greene 2007) have argued that we should distrust certain characteristically “deontological” judgments in trolley-problem scenarios because they are generated by evolutionarily-old, highly domain specific, system 1 areas of the brain such as the insula, amygdala, and ventromedial prefrontal cortex. Mark Alfano has responded that while this type of argument might be justified for the insula and amygdala (and thus, our moral intuitions based on disgust reactions should not necessarily be trusted), the ventromedial prefrontal cortex is in fact involved in processing, not generating diverse emotional and value-related inputs and thus cannot be used to debunk deontological responses to trolley-problem scenarios (Alfano 2016, 103–4). Now, this particular debate won’t be of concern to us, but the general pattern of justification is: we will see in the next chapter that one of the lines of argument in support of the Consequences and Character

⁴² One might object here that that this conflates considerations about the best moral theory to adopt and considerations about the best ways of making people act morally (thanks to Martin Montminy for raising this question). To a certain extent, it does just that: it treats moral systems as systems that need to be used, and the fact that some candidate moral system worked against all our natural tendencies would be a mark against it on this view. This is the key claim generated by the design paradigm: that these types of considerations are legitimate considerations for deciding moral theories, and systems of moral-theoretical justification that rely only on (for example) our traditional moral values are overly narrow.

conditions argues just that they better utilize some of our natural tendencies in moral evaluation than competing moral theories (without coming at an additional a cost).

Second, consider examples of good design that is good because it better directs our attention to where our attention should be directed. Most famously, artists and performers in various fields have developed design strategies for directing attention for millennia. To just mention a very, very few examples: visual artists use lighting, contrast, and the gaze of depicted characters to direct our attention; musicians use crescendos, silences, and scales; orators and actors use dramatic pauses and volume changes; authors use the white space of the page.

But for us, the paradigm example of design aimed at directing attention is the case of rumble strips on highways. When a car starts to drift out of its lane, it hits the rumble strips, and the rumble strips provide a warning by directing the driver's attention to the fact that they are drifting. This is, in the circumstances, very useful information. When we are in that type of situation, we would very much want our attention to be directed to that fact. Similarly, different moral systems can more or less effectively direct our attention to what they take be most morally important, pertinent, or salient in or about particular situations. For example, Hagop Sarkissian (2010) has argued that something like this idea was behind the Confucian emphasis on ritual and its relation to moral development. Effectively designed rituals encouraged and inculcated particular emotions that helped make salient the most important moral information in the situations for which the rituals were prescribed. This general point, that ethical theories can be better or worse designed insofar as they better or worse direct our moral attention, will also be important to us: another of the lines of argument in support of the

Consequences and Character conditions will claim that they help direct our moral attention to morally important aspects of situations.

A third nice feature of good design is that, in addition to facilitating the achievement of its intended goals, it provides further benefits or positive externalities. Consider the design of the traffic roundabout in civil engineering. The goal of a roundabout is to direct traffic, and they have existed much longer than stoplight intersections, but it was not until more recently (partially with the design of the “modern roundabout”) that additional benefits or positive externalities of roundabouts became more apparent. Because of their design, not only do they reduce collisions, but they also reduce their severity because drivers are moving more slowly and not moving in directions that allow for t-bone or head-on crashes (“Roundabout Benefits,” n.d.). But, roundabouts also cost less than stoplights, still function when the electricity goes out, and they drastically improve the flow of traffic and reduce delays, and take less overall space than intersections by requiring fewer lanes approaching the intersection to direct traffic (“Roundabout Benefits,” n.d.). Analogously, ethical theories might be *ceteris paribus* better or worse designed depending on their additional costs and benefits – the positive and negative externalities of the claims to which they are committed. A well-known example of this type of consideration is that we should believe in free will, because failing to do so would lead to bad effects (see Smilanksy 2002; Pereboom 2006). Similarly, Manuel Vargas (2013) has argued that compatibilist theories of moral responsibility that promote agency and agency cultivation should be preferred over theories of moral responsibility that don’t, because these indirect effects are desirable. Or, consider arguments in favor of religious tolerance (Locke 2010; Mill 1974), or

moral pluralism (Bommarito and King Forthcoming) that do not attempt to support these views by claiming that there actually are multiple correct religious systems or multiple correct moral systems, but instead that given human beings' propensity for out-group prejudice, xenophobia, and jingoism, we are much better off to tolerate a wide range of moral or political views. This general form of argument will also be important to us: another of the lines of argument we'll see in the next chapter in support of the Character condition claims that it provides an important additional benefit of putting extra meaning and value into our world.

At this point, our picture of perspectival naturalism is starting to be filled out, and we can begin to see how perspectival naturalism meets several of the desiderata for our metaethical theory that we set out at the beginning of the chapter. First, we wanted our view to be scientifically and philosophically defensible – especially given the radical debunking argument. So far, nothing I have said about perspectival naturalism is inconsistent with the conclusions of the radical debunking argument. The second desideratum was to provide a framework for justifying moral claims that also gave us a path forward in cases of moral disagreement without attempting to derive justification on Kantian constructivist grounds of some kind of universal convergence on a particular set of moral claims. We have now developed enough of the theory to have a general picture of how perspectival naturalism will do this. Most generally, perspectival naturalism inherits from constructivism the view that what justifies moral claims or theories *at all* is that we are valuing creatures: that we can occupy practical standpoints. But, this does not tell us which particular moral claims or theories we should adopt. Diverging from certain anti-realist theories that tell us we should adopt the moral

theories that best fulfill some particular function of morality, perspectival naturalism instead claims that we should adopt the moral theories that are best designed *for us*: those that best facilitate achieving our *various* moral goals (some of which will be mutually supporting, some of which will be in tension), given our particular values, preferences, circumstances, perspectives, patterns of behavior and thought, resources, and capacities, without over- or -underemphasizing any of these features (where what counts as over- or under-emphasizing is itself determined by our various moral goals, values, preferences, circumstances, perspectives, patterns of behavior and thought, resources and capacities).⁴³ Further, the design paradigm shows how perspectival naturalism can provide paths forward in moral disagreement without being committed to a Kantian constructivist convergence on some universal moral theory: unlike a functional account of morality, design problems cannot be solved via a single decision procedure (literally, a function) alone, but we *can* provide better or worse solutions to design problems.

We should pause here to briefly address some immediate concerns. First, consider the following questions: Back during our discussion of the radical debunking argument, I claimed that we could not gain any knowledge from objective moral facts - even the most basic knowledge such as whether objective moral facts were consistent with each other. Can perspectival naturalism do any better on this front? What justification would perspectival naturalism provide for constructing or designing

⁴³ One might worry here that the claim is circular: that the definition of moral justification includes our moral goals, values, perspectives, and capacities (thanks to Martin Montminy for raising this question). In fact, the structure is not circular but recursive: we begin with various moral inputs given to us by evolution, develop moral systems based on those inputs and constantly re-evaluate and refine that system (and the inputs themselves) based on the inputs.

consistent moral theories? Somewhat relatedly, if even consistency in our moral theories is in question, so too are central moral concepts such as rightness. But, the main thesis of this dissertation is *about* rightness. What justification would perspectival naturalism provide for constructing or designing moral theories that include *rightness* as a concept at all?

Answering these questions will be useful in two ways: first, they are good questions that need a plausible perspectival naturalist response, and second, answering them will help illustrate in a very general way the types of considerations that perspectival naturalism takes to be relevant for justifying moral systems. Now, my goal here is not to provide a full perspectival naturalist answer to these questions: doing so would require far more detail than what we strictly need for our purposes. Instead, I just want to sketch some of the types of considerations that would be relevant for a perspectival naturalist. So, why should we prefer consistency in the moral systems we develop? First, we should notice that descriptively, we often *do* care about moral consistency: we are prone to noticing and condemning moral hypocrisy. Fairness is a centrally-held moral value, and contradictions in moral principles (and thus decisions based on these principles) are prime examples of unfairness. Thus, being concerned with consistency in our moral theories utilizes natural and antecedently-developed tendencies. Further, consistency as a value has proved epistemically useful for non-moral truths, and much of our epistemically reliable thought patterns and practices rely on treating inconsistency as a sign of falsehood. Constructing moral systems that allowed for inconsistencies would have the negative externality of leading to an uncomfortable disconnect between our moral epistemic practices and non-moral

epistemic practices. Thus, though there is no reason to think that our best justified moral systems are comprised of moral facts that are objective and consistent, there are still reasons to aim for consistent moral systems as far as is possible. Next, why should we bother with a concept of rightness? Descriptively, we have and employ such a concept, and its centrality in our moral systems provides a pro tanto reason to utilize it. Within the design paradigm, rightness fruitfully can and does fulfill two key related, but distinct functions: it can provide and communicate a quick, clear, and powerful verdict about the morality of some possible action, and it can provide an all-things-considered verdict about an action. The first function is useful insofar as there are some actions that we might want to immediately and forcefully suggest doing or not doing. The second function is useful because at the end of careful and prolonged moral deliberation, we still want an outcome and suggestion of what to do or not do. Rightness and wrongness as concepts help fulfil these useful functions, and thus seem worth including in a well-designed theory.

A second, distinct immediate concern that we should stop to address here is that there is something morally problematic with the design paradigm itself: it is hubristic, or megalomaniacal, or just plain creepy. Who do we think we are to play God and go around *designing* moral systems for others, the objection might go. I should begin by pointing out that there absolutely are creepy or megalomaniacal ways of thinking about *designing* moral systems. But, the overall concern here is not well-founded. The values that guide good design also guide design processes, and insofar as we find certain ways of designing moral systems problematic, they would not actually be justified on perspectival naturalist grounds. The real question here is whether all ways of designing

moral systems are problematic. And I think the answer must be “no.” The reason for this is that the process of designing moral systems, according to perspectival naturalism, is exactly what we have always been doing when we have ever engaged in ethical discussions, debates, deliberation, or reasoning. We could not have been doing anything else: the radical debunking argument shows that we surely were not uncovering or utilizing objective moral facts. The design paradigm, we must remember, does not claim to be offering us anything radically novel: instead, it aims to map and systematize, and to provide a framework that makes more clear and plausible and justifies the ways that we *already* engage ethically with each other. We have always been working on designing better and worse ethical theories: perspectival naturalism simply encourages us to fully appreciate this fact.

3.3.4 Perspective

Up to this point, I have not said much about the perspectival portion of perspectival naturalism. Perspectival naturalism’s commitment to a perspectival account of our minds is a key feature of the view, and will be important for distinguishing it from rival metaethical theories, for meeting the final desiderata of our metaethical theory, and for the arguments in favor of the Consequences and Character Conditions. But, what is this perspectival view of the mind, how is it different from views of the mind implicit in rival theories, why should we prefer it, and what will the implications of a perspectival view of the mind be for moral theory?

The basic idea here is that many metaethical and normative theories in contemporary analytic philosophy and ethics have inherited a focus on propositional attitudes. This means that there can be a tendency when we think of our (moral) minds

to think mostly in terms of our beliefs (and perhaps, our desires) and when we think of (moral) theories, to think mostly in terms of the claims that theories make. Perspectival naturalism emphatically rejects this tendency. Instead of being concerned solely with whether a set of moral beliefs is consistent and true, we will also be concerned with how well a moral perspective is *aligned* (that is, how the various components and structures of perspectives are aligned with each other). Analogously, instead of being solely concerned with whether a moral theory makes true claims, we will also be concerned with how well a moral system (the perspectival naturalist stand-in for theory) is aligned.

To make clear what this focus on perspective adds to our moral theorizing over and above a model of the mind as a set of beliefs (or set of beliefs and desires), we'll need to say more about what perspectives are. As I mentioned previously, we do not need to start from scratch here: there has been a small but growing philosophical literature developing exactly this idea. Over and above including our beliefs, perspectives also include patterns of association and inference and they can be shaped and constrained not only by our beliefs but also by our values, emotions, attitudes, basic evaluative tendencies, and moods. Perspectives *filter* information. Wayne Riggs (2016), claims:

... the interrelations among [a perspective's] parts are such that, within the domain of the perspective, they influence how we perceive the world, which inferences we are likely to draw (perhaps especially causal or predictive inferences), where we are likely to look for evidence regarding related propositions, whether we greet related propositions as suspicious or benign, etc...a perspective is a 'take' on some part of the world or a 'construal' of it. It is more than a simple belief or even a disconnected set or list of beliefs. (Riggs 2016, 20)

Importantly, for Riggs, a perspective is not simply a set of networked beliefs, and should not be analyzed as such (2016, 21; Camp 2006 makes the same claim). Elisabeth

Camp characterizes perspectives as “on-going dispositions to structure one’s thoughts...to notice and remember certain types of features rather than others, so that those features are more *prominent* or salient in one’s intuitive thinking...[and] dispositions to treat some classes of feature as more *central* than others...” (2013, 336). Like Riggs, Camp argues that analyzing perspective solely in terms of propositions is misguided: “...one can retain a perspective while adopting or abandoning any particular propositional claim...Further, explicitly entertaining, even endorsing a certain set of propositions – including higher-order structural propositions concerning prominence and centrality – is neither necessary nor sufficient for actually ‘getting’ a perspective” (Camp 2013, 336).

The term “perspective” of course carries visual and artistic connotations, and these have been explored and developed helpfully by Bas Van Fraassen (2008) and Catherine Elgin (2010). Drawings from linear perspective in art, these authors emphasize, is *indexical* but not subjective: “Because cross-ratios are invariant across changes in orientation and origin, drawings in perspective convey objective information about constancies in the appearances items present from different views” (Elgin 2010, 442).⁴⁴ In addition to being essentially indexical, perspectives also entail occlusion and non-commitment. Some object (or information) is occluded in a perspective when another object (or other information) *blocks* it from view. If, from where I am standing, I cannot

⁴⁴ In other words, a drawing in linear perspective models objective information about the world (e.g. the distances between objects in the drawing). The cross-ratio is a mathematical function that uses geometric properties to infer this information. One implication here for our project is that we cannot infer a relativism or subjectivism directly from perspectival naturalism: the mere fact that we always occupy (moral) perspectives does not mean that we can never access objective (moral) information about the world. The anti-realism in perspectival naturalism thus does not emerge from its perspectivalism but from the radical debunking argument.

see a horse because (from my perspective) the horse is behind the barn, the horse is *occluded* (Elgin 2010, 442). Similarly, a person's beliefs and commitments might occlude information: this is, for Kuhn (though he did not use the term "occlude") a key feature of paradigms in science (Kuhn 2012) and is a fixture of political orientations. Because perspectives always filter out some information, they are also non-committal with respect to certain properties of objects or information: they don't represent that object or information as having that property one way or another (Elgin 2010, 442). For example, a drawing of a person in a hat is non-committal about whether the person is bald.

Consider a standard moral realist account of moral progress. It would occur when we acquire new true beliefs about objective moral reality. One moral theory is better than another (*ceteris paribus*) when it includes more of these true moral beliefs. But, an account like this is lacking. In a more general epistemological context, Catherine Elgin argues:

Human beings [we might think] seem to gather information in the way that squirrels gather nuts. Bit by bit, we amass data and store it away against future need. Many epistemologists and laymen take cognitive progress to consist in data gathering...[but] this conception of cognitive progress both constricts and distorts the subject...it cannot even make sense of a variety of cognitive innovations that figure in the advancement of science. (Elgin 2002, 1)

Elgin goes on to explain:

The inadequacy of the dominant [proposition / belief oriented] view stems from its indifference to the fact that cognitive progress often consists in reconfiguration -- in reorganizing a domain so that hitherto overlooked or underemphasized features, patterns, opportunities, and resources come to light. Epistemological theories that restrict their purview to justified or reliably generated true beliefs seem blind to such progress...they lack the resources and the incentive to explain how or why the reconfiguration of a domain can itself be cognitively valuable. (Elgin 2002, 2)

Of course, Elgin is not making the claim that epistemic progress *never* consists in acquiring new information or new true beliefs, but instead that the standard focus in epistemology on acquiring new true beliefs itself *occludes* the fact that this is only one type of epistemic progress.

What all this means for us is that a perspectival model of our minds will contain a great more than merely propositional attitudes (though it will of course include those as well). That is, a perspective will include, in addition to beliefs, values, desires (and other propositional attitudes): patterns of association, what we *perceive as* having certain properties (e.g. valuable, desirable, harmful, honest, disgusting) (this is sometimes referred to as “seeing-as”), which inferences we are likely to draw (including causal and predictive inferences), where we look for and what we see as evidence, how we interpret new propositions (as being likely, suspicious, benign, for example), which features of things are more prominent / salient, which are more central, whether the perspective encodes objective information accurately, what information is occluded within the perspective, and what information the perspective is committed to or not committed to representing. And the absolutely key point for a perspectival naturalist is that each these features of perspective is morally evaluable. Certain perspectives are morally better or worse than others, and not merely for the morally-relevant propositional attitudes they contain. Perspectives that occlude certain moral information are morally worse than others, perspectives that treat as central or prominent certain moral information are morally better than others: we should not aim only to have the morally best beliefs, but also the morally best perspectives.

All of this has important implications for moral theory. For a perspectival naturalist, we are not interested only in whether a set of the claims that constitute a theory is true. In fact, instead of thinking of moral theories in terms of sets of the claims they are committed to at all, perspectival naturalism encourages us to think of moral theories in perspectival terms, as something more like paradigms. And, perspectival naturalism here will apply the Kuhnian (2012) point about scientific theories: they never actually were merely sets of propositions. Moral theories implore us not only to believe that certain things are right or wrong or good or bad, but to *see* certain things as right or wrong, to take certain moral information to be more central, to be more likely to make certain inferences than others. To help emphasize this point, I will refer (and have been referring, when appropriate) to moral theories in this paradigm sense as “moral systems.” And, just as in the case of perspective, all of these features of moral systems over and above the propositional claims they are committed to are morally evaluable. One thing we should notice here is while these perspectival commitments will force us to slightly tweak the design paradigm in perspectival naturalism, they actually fit together quite nicely. After all, the design paradigm told us that well-designed moral theories managed and brought into alignment various factors: goals, values, resources, capacities, and constraints. The commitment to a perspectival view of our minds simply adds extra (non-propositional) factors that can be better or worse managed, and better or worse brought into this alignment. This point will be important in our argument for the Character condition: part of the reason to prefer the Character condition will be that it helps better align the non-purely propositional parts of our moral perspectives and systems.

And now, finally we are in a position to meet the last of our desiderata for a metaethical view by showing how perspectival naturalism justifies dislodging strongly-held moral intuitions without throwing away our moral intuitions altogether. The argument will go like this: Call a moral perspective or system “perfect” if it captures *all* moral information to the fullest degree. Our moral intuitions exist within moral perspectives and systems that *cannot* be perfect, and since no single moral perspective or system can be perfect, the moral perspective or system I adopt should always be open to revision, and because I know this, I should always leave open the possibility that even my most strongly-held intuitions might need some revision. Why should we believe that no moral perspective system can be perfect in the stated sense? Perspectival naturalism gives us at least three reasons. First, the basic evaluative tendencies with which we evaluate moral systems were given to us by evolution, and are thus highly unlikely to be perfectly consistent. No single moral system can consistently capture them all, so we must always pick and choose. But, even if we overcame this problem, we would still not be able to develop a perfect moral perspective or system. Second, there will always be tensions stemming from the design paradigm: the design paradigm aims at best managing a litany of factors, and given the complexity of this task, it is likely that there always be well-designed, competing moral systems that manage these factors slightly differently. There will likely always be trade-offs in design. But, even if we overcome this second problem, we would still be subject to the third problem, and I think there is simply no escaping it. The third problem for developing a perfect moral perspective or system just is that these are, by definition perspectival, and perspectives, by definition, filter information. There is no view from nowhere, and so there will

always be moral information *outside* our currently-occupied perspective that *would* matter to us but is unavailable, occluded, or less prominent or central or salient than it could be. No moral perspective or moral system can capture all moral information to the furthest degree, and so I know that no matter how well-designed the moral system or perspective I hold is, there will always be moral information that I consider important available if I adjusted my moral system or perspective. Thus, moral systems and perspectives are essentially imperfect, and thus, essentially revisable: I will always have a good reason to be open to revising my moral perspective or moral system because I know there will always be important moral information not captured by it.⁴⁵ Now, it does not follow from any of this that I should completely abandon my moral intuitions: doing so would be like abandoning a tool that I reasonably believe works effectively the vast, vast majority of the time. But since I should always be open to revising my moral perspective or system, I should always be open to revising my most strongly held moral intuitions within that system (even if this amounts to shifting my credence in them down very slightly). And this is the last of what we wanted to show in this chapter. Next, we will develop the argument for the Consequences and Character conditions.

3.4 Summary

I began this chapter by developing a more radical version of familiar evolutionary debunking arguments against moral realism. According to this argument, once we truly appreciate the distorting influence of evolution, we should be extremely skeptical that we could ever discover fully mind-independent objective moral truths that we could

⁴⁵ One might worry here that this point assumes the existence of perspective-independent moral facts, which I deny (thanks to Martin Montminy for raising this concern). However, this relevant moral information does not exist independently of all moral perspectives: it exists in other (real or counterfactual) moral perspectives.

apply – even apparently self-evident moral truths. Yet, we are valuing creatures. The combination of these facts, I argued, should push us in the direction of Humean Constructivism as the most plausible metaethical view, and I made the case for a version of Humean Constructivism that I called perspectival naturalism. The main theoretical commitments of perspectival naturalism are (a) naturalism / descriptive adequacy: moral theorizing should be informed by our best descriptive understanding of the world; this understanding will be naturalistic; (b) anti-realism: there are no *knowable* objective, fully mind-independent moral truths; (c) non-nihilism / constructivism: anti-realism does not entail the abandonment of morality; moral systems are ultimately justified not by being objectively true but via our moral standpoints, (d) perspectivalism: an adequate model of the mind must treat the mind as perspectival; perspectives do not reduce without loss to sets of beliefs, and (e) the design paradigm: The best moral systems for us are the moral systems that are best designed for us. Finally, I argued that given the radical debunking argument and perspectival naturalism, we need to re-examine even many of our strongest moral intuitions about rightness and wrongness and that the best-designed moral theory for us might not be entirely intuitive to some of us right now: its justification might come from elsewhere.

Chapter 4: The Consequences and Character Conditions

4.1 Introduction

In the previous chapter, I made the case for a form of metaethical constructivism that I called perspectival naturalism. According to perspectival naturalism, there are no objective, fully mind-independent moral facts. However, some moral claims and systems can be more or less justified than others. On the account I developed, systems derive their justification from how well they are designed for us: how well they align and manage our moral goals, values, wants, needs, abilities, perspectives, and circumstances. In this chapter, I will make a perspectival naturalist case for the Consequences and Character Conditions.

4.2 The Consequences Condition

We should begin here with a reminder of the Consequences Condition:

The Consequences Condition: for at least some (but not all) cases, the rightness or wrongness of an act should be determined conclusively by the act's consequences (i.e. how well the act promotes well-being): though there may be competing moral considerations, no competing moral considerations could outweigh the consequences in determining the rightness or wrongness of the act.

In Chapter 2, I described the Consequences Condition in more detail, and in this section, I will lay out the argument for it.

We should begin the argument by mapping out the logical competitors to the Consequences Condition. On the one extreme, one might claim that the consequences of an action should *always* fully determine its rightness or wrongness. Call this the Always-Consequences Condition. On the other hand, one might claim that the consequences of an action alone should *never* fully determine its rightness or wrongness. Call this the Never-Consequences Condition. Our first order of business is

defending the basic version of Consequences Condition against these rivals. The basic structure of the argument will be this:

1. We should prefer (*ceteris paribus*) moral systems that are better designed for us.
2. Moral systems that meet the Consequences Condition are better designed for us than moral systems that meet the Only-Consequences or Not-Consequences Condition.
3. Therefore, we should prefer (*ceteris paribus*) moral systems that meet the Consequences Condition.

Since premise 1 just is the theory of evaluation held by perspectival naturalism, which I argued for in the last chapter, the heavy lifting here will need to occur in premise 2.

Let's begin with the argument against the Never-Consequences Condition.

4.2.1 The Never-Consequences Condition

Some might find the Never-Consequences Condition, according to which an act's consequences can never conclusively determine an act's rightness or wrongness fairly implausible to begin with. Consider the following cases:

Uncontroversial Evil Scientist Case: An evil scientist kidnaps you and puts you in a room with two buttons. If you push the red button, the evil scientist says, you will be released, but every other person on the planet will be tortured then killed. If you press the green button, you will also be released, but every other person on the planet will be spared, and the scientist will not repeat this experiment ever again. The scientist is telling the truth.

Uncontroversial Trolley Case: A runaway trolley is hurtling down the tracks towards a group of 400 surprisingly-densely packed and thoroughly oblivious picnickers. You are standing by a lever, and if you pull it, the trolley will switch tracks. The other track is empty and long enough for the trolley to safely come to a stop without causing any injuries.

In these cases, the right and wrong options seem quite obvious. And, any other reason for believing that the wrong options are wrong than because of the damage they do will, for many (though certainly not all) seem superfluous, redundant, and derivative. And, even many of those who emphasize the moral importance of non-consequentialist features of acts would be happy to grant all of this because many non-consequentialist

theories of rightness are aligned more with a rejection of the Always-Consequences Condition than with acceptance of the Never-Consequences Condition. Recall Arpaly's (2003) discussion of a Kantian-inspired distinction between moral desirability and moral worth:

The moral worth of an action is the extent to which the agent deserves moral praise or blame for performing the action, the extent to which the action speaks well of the agent... Obviously, the extent to which an agent deserves praise or blame for her action depends substantially on the action's moral desirability: whether it is right or wrong, or how grave a wrong it is, or whether it is the best possible action, but two actions which are equal in moral desirability may be of different moral worth. (Arpaly 2003, 69)

Even Kant, Arpaly claims, would not necessarily deny this intuition about rightness:

Recall Kant's Prudent Grocer, who prices his merchandise fairly because a reputation for honesty tends to increase his profits. Despite the Prudent Grocer's unimpressive motive, Kant never denies that the grocer does the right thing or that he performs the action required of him by duty. In this sense, Kantians clearly care about results. (Arpaly 2003, 70)⁴⁶

But, strong intuitions alone cannot decide the case. For a perspectival naturalist, strong intuitions can only ever provide part of and a starting point for the ultimate justification for a moral claim or system. In any case, even *if* the intuition that we should be able to determine, in at least some cases, the rightness or wrongness of an action solely by its consequences is quite widespread, it is not universally held among philosophers or non-philosophers. Thus, we should return to the idea of what a theory of rightness is and should be *designed* to do. Recall from the previous chapter that one of the key useful

⁴⁶ Of course, Arpaly is not claiming here that we should interpret Kant as a consequentialist. Whether we agree with Arpaly's interpretation of Kant here, her point seems to be that even a Kantian could treat rightness as a purely consequentialist concept so long as they emphasize another form of moral evaluation that includes non-purely consequentialist features. I should also point out here that this distinction between moral worth and moral desirability, as drawn by Arpaly, is at odds with my overall argument insofar as I claim that rightness should not be treated as a purely consequentialist notion.

functions of the concept of rightness is to provide and communicate a quick, clear, emphatic, and powerful verdict about the morality of some actual or possible action. One of the main reasons that a moral system with a concept like rightness is relatively better designed than a moral system without one is that it provides us with this tool for communicating when the moral stakes are extremely high. And, if one values the well-being of others *at all*, then there are both hypothetical and actual situations in which there is such a massive difference in the (welfare-related) consequences of different options that we are much better served by a moral system that readily can capture and communicate this fact.

Now, one might worry that this particular argument cannot be very convincing to those it is aimed at. After all, it is explicitly built on the assumption that consequences for well-being matter *at all*: it states that if they do matter, experience suggests that situations occur in which the goodness of the consequences of various options differ so vastly that they would swamp other moral considerations. If one does not believe that consequences for well-being matter *at all*, then the argument just provided simply does not apply. Call this claim, that consequences should not matter *at all* for rightness the Strong Never-Consequences Condition. Now, I cannot here provide a definitive argument that consequences *must* matter for judgments of rightness and wrongness. I can however sketch how a plausible perspectival naturalist argument might go. The first step of the argument would be to re-emphasize a key point of the last chapter: that our moral systems should always be open to revision. Thus, an argument for the Strong Never-Consequences Condition based solely or primarily on the strong intuition that it is true (especially, when that intuition is so far from universally shared)

is immediately problematic. We need to think about the design of the theory. And, there is an argument to be made that a theory of rightness that counts consequences is better designed based on a key design idea we saw last chapter: desire paths.

Desire paths are pathways that humans or animals naturally take to move through spaces, and good landscape design will attempt to utilize them to the degree that it can without interfering with other goals. In perspectival naturalism, the analogue for desire paths are our natural moral tendencies. Sometimes we do need to utterly reject them, especially when they clash with our other important moral considerations, but otherwise it is worthwhile to try to utilize and steer them than to work directly against them. There is evidence from moral psychology to suggest that moral concern about consequences is just this type of natural tendency. The first, and perhaps unsurprising step to showing this is to emphasize the (as far as I know) unanimous agreement in empirical moral psychology that judgments of harm affect moral judgments: so far as I know, no one in empirical moral psychology has claimed or provided evidence that judgments of harm *never* influence moral judgments. In fact, nearly the opposite view has been much more popular: that judgments of harm are among the most, if not *the* most influential factors on moral judgments. When people provide moral condemnation of something (comic books, violence in films or video games, pornography, the war on drugs, trigger warnings, border family separations of asylum seekers, acts like murder, assault, bullying, or stealing, the rights of homosexual people to marry and raise children) their reasons often include claims that the item or behavior in question is harmful in some way, and often the debate about whether what has been condemned is actually deserving of moral condemnation revolves around

whether or not it actually *is* harmful (this is emphasized in Schein and Gray 2017). Early developmental psychologists found that recognition of harm was used by children to separate conventional from moral transgressions (Huebner, Lee, and Hauser 2010; Piaget 1965; Turiel, Killen, and Helwig 1987). Both infants as young as 6 months old (Hamlin 2012; Van de Vondervoort and Hamlin 2016) and many non-human primates (Sheskin and Santos 2012) track harm in social evaluations. There is a simple evolutionary story here: harm straightforwardly decreases the likelihood of survival and reproduction, so we evolved to track it (Haidt and Joseph 2007; Hauser 2006). Certain harmful acts seem universally condemned (Mead 1961; Mikhail 2007, 2009; Shweder 2012). Further, research in empirical moral psychology has found that harm seems to be the most frequent concern in daily moral judgments among both liberals and conservatives (Hofmann et al. 2014), and when people are asked to think of an example of something that is “morally wrong” over 90% offer an example of a harm such as murder or abuse (Schein and Gray 2015).

But, there is another layer to the argument here. Kurt Gray has hypothesized and argued for the existence of a pattern of moral reasoning he calls the “dyadic loop” (Gray and Schein 2016; Gray, Schein, and Ward 2014; Gray, Waytz, and Young 2012; Schein and Gray 2017). The dyadic loop treats our judgments of harm and our general moral judgments as bidirectional and mutually reinforcing: if I judge something to be harmful, I’ll much more easily judge that it is wrong, and if I judge that it is wrong, I’ll much more easily judge that it is harmful (Schein and Gray 2017, 2; see Gray and Schein 2016). Through personal deliberation and social discussion, acts become “harmified” and “wrongified” (Schein and Gray 2017, 2). As we think or talk through the morality

of an act, we will naturally gravitate to considerations of harm, and these considerations will in turn influence and reinforce our judgments about the morality of the act.⁴⁷ Much of the evidence for the dyadic loop in our moral reasoning comes from paying attention to actual moral and political discussions and debates: when new laws are discussed or passed, lawmakers often attempt to justify them by claiming there is some harm they are trying to eliminate: Schein and Gray (2015) mention, for example, political arguments about censorship, same-sex marriage, and bathroom use. Much of the evidential support from the lab for the existence of the dyadic loop comes from work showing the *facility* with which judgments of harm co-occur with moral judgments. For example, judgments of harm predict the severity of moral judgments and the perception of harmfulness is a better predictor of immorality ratings than disgustingness or weirdness (Schein and Gray 2015). One study (Schein, Goranson, and Gray 2015) asked participants to categorize 60 different acts on immorality, harmfulness, and unpleasantness and found that the more harmful the act, the more *quickly* it was categorized as immoral. One of the most fascinating examples of the dyadic loop in action is from (Rottman, Young, and Kelemen 2017), who found that children would accept testimony that something was wrong without being given harm- or welfare-based reasons (e.g. given disgust based testimony from adults) but then later on, when asked why that thing was wrong, would provide harm- and welfare-based reasons.

⁴⁷ We need to be careful here. Gray's work is most well-known for his development of the Theory of Dyadic Morality (TDM) according to which *all* of our intuitive moral judgments are based on a cognitive / category template of an intentional agent causing harm to a vulnerable patient (Gray and Schein 2016; Gray, Waytz, and Young 2012; Schein and Gray 2017). The argument here does not treat TDM in general as true: the existence of the dyadic loop does not depend upon the truth of TDM.

And so, the argument is this: just as landscape engineers should, absent strong countervailing reasons, utilize rather than work against desire paths, so too should we utilize our natural evaluative tendencies rather than work against them (absent strong countervailing reasons). Our natural evaluative tendencies are especially prone not only to noticing welfare-based considerations in the first place, but as we reason and deliberate about moral issues, to further noticing and appealing to welfare-based considerations. And so, since it would be doubly difficult to stave off these tendencies, we would need strong countervailing reasons to want our moral system do so. Absent such strong countervailing reasons, we have a strong *pro tanto* reason to reject the Strong Never-Consequences Condition.

Before we move on, there is a separate, subtler worry we should briefly address here which relies on the second key function of the concept of rightness (that it provides an all-things-considered verdict about some action): that in subsuming the rightness of an action solely to its consequences for well-being we are liable to (and in many cases will) neglect other important moral information. There is, after all, more to our moral lives than can be counted in consequences alone. However, this intuition is compatible with the Consequences Condition. After all, the Consequences Condition does not state that the consequences *alone* determine rightness in all cases. It does not even state that consequences are the only domain of moral consideration that can *conclusively* determine the rightness or wrongness of a moral action. It is entirely compatible with the Consequences Condition that say, the cruel motivations of some possible action help determine or conclusively determine that it is wrong.

4.2.2 *The Always-Consequences Condition*

The Always-Consequences Condition states that the consequences of an action always conclusively determine its rightness – no other moral consideration comes into play. This will, for some, be more intuitive than the Not-Consequences Condition.

Consider Mill:

The motive has nothing to do with the morality of the action, though much with the worth of the agent...The motive, that is, the feeling which makes him will so to do, if it makes no difference in the act, makes none in the morality [of the act]; though it makes a great difference in our moral estimation of the agent. (Mill 1979, 17–18)

And Arpaly:

it is perfectly consistent to view the moral desirability [rightness / wrongness] of actions as depending entirely on their expected consequences, and the moral worth of individual actions as depending to some degree on the agent's motives (Arpaly 2003, 70)

The thought here is that the function of the concept of rightness *just is* to pick out the actions with the most desirable consequences, or the consequences that best promote well-being. But, this will not do. We saw in Chapter 3 that the radical debunking argument casts doubt on all “self-evident” moral claims such as “rightness just is a measure of consequences.” Self-evidence alone cannot justify our adoption of a moral claim or moral system in much the same way that self-evidence alone cannot justify our adoption of a particular mathematical claim or system: since there are mathematical systems with self-evident truths (in those systems) that seem to have nothing to do with modeling anything in our world, we always need further evidence that self-evident claims are part of a system that actually models our world. But further, it is not at all clear that a claim like “rightness just is a measure of consequences” *is* self-evident or

intuitive to anyone other than certain philosophers: people are often moved by moral considerations other than direct harm or benefit to well-being.⁴⁸

Thus, a perspectival naturalist should treat as open the question of whether rightness just is a consequentialist concept and investigate whether such a theory is better designed than its alternatives. And, it seems to me that it is not, for several reasons. First, we should note and appreciate the much greater flexibility of the Consequences Condition than the Always-Consequences Condition. In Chapter 3, I argued that perspectival naturalism motivates an open-mindedness and pluralism towards ethical theorizing: since we are never justified in thinking we have a single, perfect moral system, we should always be open to the existence of other moral systems that might be just as well-designed. This commitment to pluralism will suggest slight suspicion of narrower moral theories and slight preference for broader moral theories unless that narrowness can be justified or that broadness should be called into question. But, it seems that supporters of the Always-Consequences Condition attempting to justify their narrower view of moral considerations are in an evidential bind here: every single case that could be used to push intuitions in favor of the Always-Consequences Condition *also* supports the Consequences Condition. That is, any case that, upon consideration and reflection seems to support the Always-Consequences Condition can be captured by the Consequences Condition since the Consequences Condition grants

⁴⁸ This has been a subject of heated debate in moral psychology, where Kurt Gray (Gray and Keeney 2015; Gray and Schein 2016; Gray, Schein, and Ward 2014) has argued that all *intuitive* moral wrongs related closely with harm, while others (Alicke 2012; Haidt, Graham, and Ditto 2015; Pizarro, Tannebaum, and Uhlmann 2012) have vehemently disagreed. However, even Gray does not argue that harm is the *only* moral consideration in actual act evaluations, and, vitally, the intentions of the agent causing harm are an essential component of Gray's model of our cognitive template of harm and moral judgment (Schein and Gray 2017).

that there are cases in which the consequences alone *should* fully determine the rightness of the act. And so, if a supporter of the Always-Consequences Condition wants to provide a definitive argument that will support the Always-Consequences Condition *over* the Consequences Condition, they should show that in *no case* should non-consequentialist moral considerations matter for our judgments of rightness. And, that type of negative argument is never easy to make successfully. If we can identify even one case in which non-consequentialist considerations matter for rightness, then we should reject the Always-Consequences Condition. And, that is my exact project in the next section of the dissertation, where I provide the argument in support of the Character Condition.

4.2.3 Summing Up: The Argument for the Consequences Condition

Thus far, the argument in favor of preferring the Consequences Condition has been this: logically, we should prefer one of the Never-Consequences, Always-Consequences, or the Consequences Condition, and there are strong reasons to prefer the Consequences Condition over either the Never-Consequences or the Always-Consequences Conditions. But, it will be helpful here to shift focus away from the negative cases against the Always- and Never-Consequences Conditions and towards the positive case for the Consequences Condition. After all, it might be the case that none of the options is preferable, and I have merely focused on the problems of two of them while ignoring the third. So, what are the attractive features of the Consequences Condition? We have actually already seen them, but it will be helpful here to re-focus on them.

The most attractive feature of the Consequences Condition is its flexibility. It captures a great deal of our moral intuitions about the importance of people's welfare without being radically reductionist. Thus, it is consistent with many normative theories, including most versions of consequentialism, some versions of deontology (e.g. that of Ross), and agent-based theories of virtue ethics, while allowing space for non-consequentialist moral considerations. It builds on natural evaluative tendencies to easily notice harm both immediately and to easily appeal to welfare-based considerations in moral deliberation and discussion. Finally, in being a Condition focused on *conclusive* determinations of rightness and wrongness, it helps utilize one of the main useful features of the concepts of right and wrong: they can provide quick, emphatic communication about the desirability or undesirability of certain courses of options when there is great moral urgency.

4.3 The Character Condition

We now turn to the Character Condition:

The Character Condition: in some (but not all) cases, the moral character of a person should influence the determination of the rightness or wrongness of that person's acts.

The argument here follows the overall pattern we saw in the argument for the

Consequences Condition:

1. We should prefer (*ceteris paribus*) moral systems that are better designed for us.
2. Moral systems that meet the Character Condition are better designed for us than moral systems that meet the Only-Character or Not-Character Condition.
3. Therefore, we should prefer (*ceteris paribus*) moral systems that meet the Character Condition.

The Only-Character Condition states that character is the *only* moral consideration that should influence our judgments of rightness and wrongness. Fortunately, here I do not

owe any extended argument against the Only-Character Condition for the simple reason that I have already given it. The Consequences Condition is straightforwardly inconsistent with the Only-Character Condition – if the consequences alone can sometimes conclusively determine the rightness or wrongness of an action, then the Only-Character Condition should not be accepted. So long as the argument in favor of the Consequences Condition is convincing, we should not adopt the Only-Character Condition. So, we will need to focus instead on the Not-Character Condition, which states that the character of a person performing an act should *never* influence our judgment of the rightness or wrongness of that act. So, we will try to show that a moral system that meets the Character Condition is, *ceteris paribus*, better designed than one which meets the Not-Character Condition: that is, that the Character Condition better helps us manage and align our goals, values, commitments, principles, wants, needs, abilities, circumstances, and perspectives. To make this case, we cannot rely on a single line of argument, but instead show how the Character Condition is better designed along *several* trajectories than the Not-Character Condition.

4.3.1 Person-centered accounts of moral judgment and the practical perspective

The first line of argument runs parallel to the argument we saw against the Not-Consequences Condition: that it functions as a “desire path” for our natural evaluative tendencies. Just as landscape or architectural designers should, unless there are other overriding design considerations, attempt to utilize, manage, and steer desire paths rather than work against them, so too should moral systems attempt to utilize, manage, and steer our natural evaluative tendencies (unless there are other overriding moral considerations). This directs us towards two immediate questions: first, whether the

Character Condition better utilizes our natural evaluative tendencies, and second, whether there are other overriding moral considerations. I will address the first question directly in this section, and I address the second indirectly in the following sections when I describe some of the other moral considerations in favor of the Character Condition.

According to supporters of “person-centered accounts” of moral judgment (Uhlmann, Pizarro, and Diermeier 2015; Pizarro and Tannebaum 2012), researchers in empirical moral psychology have often been influenced and are in some cases are explicitly committed to purely act-focused normative frameworks (Uhlmann, Pizarro, and Diermeier 2015, 73). Early influential researchers in moral psychology such as Piaget (1965) and Kohlberg (1969) leaned heavily on deontological approaches to ethics, while several more recent researchers have often been committed (though to varying degrees) to consequentialist frameworks (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; see, for example, Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Uhlmann et al., 2015, p. 73). Though consequentialist and deontological traditions in normative ethics are far from silent on issues of moral character, their focus has often been directed towards acts and their permissibility, rightness, and goodness. This focus, supporters of person-centered accounts claim, occludes key information about the nature of moral judgments we make: specifically, that we tend to be very interested in making judgments about a person’s moral character and that these judgments influence our other moral judgments.

The person-centered account of moral judgment, on my reading, is committed to the following four claims related to moral character and moral judgment (Uhlmann, Pizarro, and Diermeier 2015, 73):

1. Individuals have motivations to morally evaluate character, and not just the (standardly consequentialist or deontological) rightness or wrongness of an act.
2. Acts are perceived as more or less informative of an individual's moral character, and those acts that are perceived as especially informative of moral character are weighed more heavily in moral judgments of acts.
3. Moral evaluations of acts and of character can diverge and are thus empirically distinguishable.
4. Moral evaluations of character can influence moral judgments other than (standardly consequentialist or deontological) judgments of rightness or wrongness (e.g. judgments of intentionality, agency, and blame)

Defenders of person-centered accounts have not, to my knowledge, provided an explicit and detailed account of what they mean by “character.” Most often they talk about character in terms of personality traits such as integrity, empathy, trustworthiness, fair treatment of others, or more generally, global character traits such as whether a person is a “good” or “bad” person (Uhlmann, Pizarro, and Diermeier 2015, 74; Pizarro and Tannebaum 2012, 92–93). At times they also include a person's perspectives, deeply-held moral principles and values as part of a person-centered account – they claim moral judgments track features of actions that are “informative in that they indicate who the actor is and what he or she values and considers when performing moral actions” (Pizarro and Tannebaum 2012, 104; Uhlmann, Pizarro, and Diermeier 2015, 77). On my reading, by “character,” supporters of person-centered accounts of moral judgment are thinking of morally relevant traits of varying degrees of generality, long-term and deeply-held moral commitments such as values and principles, and, potentially, the moral perspectives and patterns of moral reasoning and inference that tend to be adopted.

Defenders of person-centered accounts not only claim that we have motivations to evaluate the character of others, but these motivations run deep. There is growing evidence in the broader psychological literature that morality is an important aspect of the way people judge each other, even in brief encounters when not much information is available (Brambilla & Leach, 2014; Goodwin, 2015). Person perception researchers have traditionally found that new people tend to be judged along two main dimensions: warmth and competence (see Fiske, Cuddy, and Glick 2007). Evidence suggests that people make inferences about a person's character "spontaneously and automatically, that the ability to make character evaluations arises at surprisingly early stages of moral development, and despite some cross-cultural variability, character evaluations are widespread across cultures" (Uhlmann, Pizarro, and Diermeier 2015, 74; see Choi and Nisbett 1998; Fiske, Cuddy, and Glick 2007; Hamlin, Wynn, and Bloom 2007).

Uhlmann, Pizarro, and Diermeier (2015, 74) suggest that these evaluations might be prudentially useful. Accurate evaluations of trustworthiness help predict whether or not a person will defect in future cooperative endeavors, accurate evaluations of unfair treatment suggest that a person will not share resources equitably, and accurate evaluations of a person as lacking in care, concern, or empathy suggest that a person will not be motivated towards pro-social action in the future (Uhlmann, Pizarro, and Diermeier 2015, 74). Thus, they claim that the ability and the desire to judge the character of others via their acts might be a feature of our moral psychology that has been affected by natural selection (Pizarro and Tannebaum 2012, 93). Miller (2007) has also suggested that the ability and motivation to judge character by a person's acts

also improves long-term mate selection and so may have also been influenced by sexual selection.

According to person-centered accounts of moral judgment, actions that are taken to be particularly informative of a person's character will be weighed more heavily in moral judgements of particular acts. Research has shown many areas in which certain acts are taken to be more informative of a person's overall character. For example, negative acts are seen as more informative than pro-social acts, likely because people might be motivated for pro-social acts by prudential or self-serving reasons (e.g. to improve one's reputation) (Reeder and Brewer 1979; Uhlmann, Pizarro, and Diermeier 2015, 74). In general, acts whose intentions and motivations are ambiguous tend to be treated as providing less information about a person's character (Snyder et al. 1979), while decisions that are taken easily and quickly (Critcher, Inbar, and Pizarro 2013; Verplaetse, Vanneste, and Braeckman 2007), that are accompanied by genuine emotions (Trivers 1971), and that carry costs for the decision maker (Ohtsubo and Watanabe 2008) are taken to provide more information about a person's character. All this taken together suggests an argument for the Character Condition from the design perspective: one useful function of moral judgment is identifying people who are most and least likely to be competent cooperators in the future, and naturally judgments of rightness which are compatible with the Character Condition track this information and thus provide an additional useful function.

Stepping back, the overall argument here is this: the traditional philosophical distinction between act evaluation and person evaluation is likely not so clear cut in our moral minds: we have strong natural evaluative tendencies to run them together. Our

evaluations of a person's character will tend to influence our evaluations of their acts. Now, according to perspectival naturalism, this provides us with a (defeasible) reason to, instead of insisting we keep these types of evaluations distinct, which would be psychologically difficult if not impossible, to try to understand if and how these tendencies can be managed and steered and to steer them in ways that best align with the totality of our moral considerations. To undertake that project, we would need to show that some way of doing this can actually capture our values, and showing that will be our goal in the next two sections.

4.3.2 Goodness in actual and possible consequences

In the previous section, I argued that the Character Condition better utilizes our natural evaluative tendencies than the No-Character Condition. But of course, this alone cannot suffice as an argument for the Character Condition. There are a great many other moral considerations to weigh before we should think that a theory of rightness that meets the Character Condition is *ceteris paribus* better designed than one which doesn't. Most important among these considerations is how well it captures and aligns our considered moral values, and our focus in the next two sections is to develop the case that it does. We'll begin by looking at and adopting a recent line of argument made by Phillip Pettit (2015).

Up until now, when I have talked about consequences I have done so in a fairly typical fashion: as the actual direct and indirect consequences an action has on people's wellbeing. But recently, Philip Pettit (2015) has made the case for a more nuanced understanding of the goodness of the consequences of an action. Pettit begins his argument by drawing our attention to three of the most important goods in life (beyond

basic necessities): attachment (e.g. love, friendship), virtue (e.g. kindness, compassion, courage, honesty, open-mindedness), and respect (Philip Pettit 2015, 1). To provide an adequate account of the goodness of the benefits these provide to us, Pettit argues, we must understand them as *robust*, rather than thin goods or benefits. Robust benefits, according to Pettit, include not only the immediate and tangible benefits that you provide another when you, for example, perform an act of caring, or truth-telling, but also the counterfactual benefits you would provide in other relevant possible scenarios. To motivate and support this argument, it'll help to look at some examples.

Pettit begins explaining the robust good of attachment with the example of Gwendolen, from Oscar Wilde's comedy *The Importance of Being Earnest* (Philip Pettit 2015, 11). Gwendolen falls in love with "Ernest," explaining, "My ideal has always been to love some one of the name Ernest. There is something in that name that inspires absolute confidence" and later that "The moment [my cousin] first mentioned to me that he had a friend called Ernest, I knew I was destined to love you." The problem is that "Ernest" is only an assumed pseudonym: the character's real name is Jack. Intuitively, it seems that Gwendolen is not really in love with Jack *or* Ernest: only the smallest change in circumstances (an accurately reported first name) would drastically change her opinion of and behavior towards him. Pettit expands on this intuition:

But under our received conception of love the care you offer me in this way must not be premised on exactly how I happen to be: it must be able to survive a variety of possible changes in me, among them the change or apparent change in the name I bear. If you love me, so the lesson goes, not only must you feel and offer care that would survive sickness as well as health, poverty as well as affluence, to cite the standard vows. You must also feel and offer me care independently of how I currently look, what I currently do, or how I am currently called. Shakespeare already made the point in Sonnet 116: 'Love is not love, Which alters when it alteration finds'. (Philip Pettit 2015, 12)

Pettit thus distinguishes between care and love. The idea, on this view, is that care corresponds only to the actual benefits provided in a situation, and I can provide care for someone without loving them (for example, out of pure self-interest). To genuinely love them, I would not only provide care under the actual circumstances, but also under other relevant (some likely, some unlikely) counterfactual circumstances:

In order for me to enjoy care at your hands, you have to make things thus and so only in the actual circumstances in which we interact. In order for me to enjoy love at your hands, you also have to make things thus and so in a range of possible scenarios, some of them very improbable. While actual care itself requires nothing counterfactual, actual love requires counterfactual as well as actual care.

Pettit then goes on to note that many virtues share a similar structure. For example, the robust good of honesty requires more than the thin-good of truth-telling (Philip Pettit 2015, 46–47). The (thin) benefit in the actual world of you being honest to me is that I get true information from you (or at least, your true opinion). But, intuitively, honesty requires more than just truth-telling in the actual world. If you are only telling the truth now to gain my confidence and trick me later, you aren't being honest. If you are only telling the truth because you are being compelled to (for example, under penalty of perjury), you aren't honest. Of course, one might respond here that all this shows is that sometimes we value good motivations in addition to good consequences, and so this discussion does not actually provide any support for the Character Condition. But, importantly, even cases in which a person has a good motivation can fall short of being affection or virtue, on this view, if that good motivation would not hold across a variety of relevant counterfactual situations. It is perfectly possible for me to perform an act with good consequences and good motivations that is nonetheless out of character for me, and Pettit's point is that the fact that certain kinds of acts are *in character* for a

person is in fact a key part of what we cherish not only about those persons but also about those acts.

And so, for Pettit the key difference between offering robust goods or benefits to others and thin goods or benefits lies in our *dispositions* to provide these robust goods:

When you give me favour out of friendship, the disposition of friendship does not just trigger the production of that act and, with its work done, leave the scene. The presence of the disposition means that you produced that favour as part of a larger project of controlling for the production of favour across any of a certain range of possibilities. (Philip Pettit 2015, 4)

What we actually value most, what we in fact “cherish” is not the thin benefits that attachment or virtue provide, but the disposition to provide robust benefits (Philip Pettit 2015, 23–24).⁴⁹ And this is no small or unimportant matter: as Pettit rightly claims, these are among the things (other than basic necessities) that we value the most in life. And so, we might think, any moral theory that did not capture this value is inadequate.

At this point, a supporter of the No-Character Condition might exclaim that I have just undermined my own argument. After all, here we can develop an alternative argument that would capture every intuition I could bring to bear that character matters for some judgments of rightness and wrongness, but do so in purely (welfarist) consequentialist terms: they just happen to include counterfactual consequences in addition to more familiar actual consequences. But, there are at least two issues with

⁴⁹ Of course, one might object that we don’t in fact value or cherish the *dispositions* to provide robust goods, but in fact something else. For example, one might object that I don’t really cherish the disposition to provide robust goods, but instead I simply value the disposition because it comes with an increased likelihood of receiving thin benefits (Philip Pettit 2015, 111–15). But, this seems psychologically implausible: when I value or cherish a person’s character and dispositions to act in certain ways, it is not because I think they can provide me with the most benefits but that I think those dispositions genuinely reflect goodness in them.

this move. First, and less importantly, it is not clear that all consequentialists would or should accept this move. After all, I take it to be a key motivating factor for many consequentialists that their moral theories track *actual suffering* and *actual happiness* occurring in the world that competitor theories can be forced to ignore or underappreciate in lieu of other moral considerations. Think here of act utilitarian criticisms of rule utilitarianism (e.g. Smart 1956): part of the problem with rules that *would* generally maximize total utility is that in some instances they would not, and this should bother people whose initial or primary moral concerns were focused on human happiness and suffering. Now, of course, this response won't be convincing for all consequentialists, and especially not the kind of consequentialist who is most likely to put forward this criticism in the first place. But, there are more concerns with this line of argument. The second is that it rests on a simple ontological mistake (a mistake that it is worth pointing out Pettit does not seem to make in his discussion of robust goods): the counterfactual consequences are grounded in our actual dispositions of character and not the other way around. We should not value or cherish the ways a person would act in the types of situations Pettit has in mind without valuing or cherishing the dispositions of character which would cause them. Third, for this line of criticism to go through, it would need to be the case that the constraints on which counterfactual situations are *relevant* can themselves be cashed out in fully consequentialist terms. But, it is unlikely that this could succeed. In some of the counterfactual situations we are interested in, the act of affection or virtue very well might lead to slightly worse overall consequences because of bad moral luck, and in others the act might not actually be possible to complete or even might not have any effect on the consequences at all. For

example, in at least some of the relevant counterfactuals for honesty are situations in which the person *would* act honestly even when (unknowingly to them) the exact same result would happen whether they were honest or not. But, if these counterfactuals are relevant, it is difficult to see how they could be so on purely consequentialist grounds. Finally, and most importantly, there is a fourth issue with this line of argument that seems decisive to me: we do not value or cherish *only* the counterfactual consequences of the types of acts in question. We also cherish the motives, emotions and commitments that the people hold – both the occurrent and the dispositional ones. We do this even in cases where those motives, emotions, and commitments would not succeed in actually improving our or anyone else’s welfare: sometimes it really is the thought that counts. So, if we accept that our theories of rightness should track what we value and cherish most about acts (which was the foundation for being interested in counterfactual consequences in the first place), we should not think that we can do so by appealing *only* to the counterfactual consequences of acts. If we are basing our account of rightness on what we value – and not merely the consequences of what we value – it seems we must be in for a penny, in for a pound here.

A well-designed theory of rightness should capture what we ought to value or disvalue about acts. And, if the claims in the discussion we’ve just been having hold, then, for some (but not all) acts, we value more than only their immediate and actual consequences and even their immediate and actual motivations.⁵⁰ A theory of rightness

⁵⁰ Importantly, for Pettit, it is not the case that all acts should be evaluated partially on the basis of the character of the person who performed them. He distinguishes between act-, action-, and agent-evaluations. Act- and agent- evaluations are familiar: they correspond roughly to the traditional distinction we’ve already seen between judgments of moral rightness and judgments of moral praiseworthiness. Action-evaluations,

that only captured those would be, for some cases, simply incomplete. And so, we have borrowed from Pettit a straightforward argument in favor of the Character Condition. But, there are other ways that we value character's influence on our actions and other reasons we have to build this into our theory of rightness.

4.3.3 *Individual Differences in Character*

In this section, I will present two further lines of argument in favor of adopting the Character Condition. We'll explore two ways in which this happens. The first is that sometimes part of what we value about an act was that it was performed out of the deeply-personally held commitments, values, and perspectives of the person who performed it: that it was an act of integrity. The second is that sometimes part of what we value about an act was that it was performed out of the *particular* commitments, values, and perspectives of the person who performed it.⁵¹ The general idea for both lines of argument is this: our theory of rightness should capture what we value about acts, and part of what we value about some acts is that they were performed out of deeply-held commitments, values, and perspectives that we in turn value or cherish or that they were performed in certain particular ways or from particular perspectives that we value or cherish. Further, according to perspectival naturalism, there are many values or commitments that we could and would hold if we were to simply shift perspective: we can (and often do) appreciate a far greater moral complexity than we

however capture the robust goods of actions, which essentially involve the dispositions out of which they are performed (Philip Pettit 2015, 4). This distinction, it is worth noting, is perfectly consistent with the Character and Consequences Condition, because for Pettit, rightness and wrongness apply to both act- and action- evaluations. And since we are much more interested in rightness and wrongness here than action theory, I will be agnostic towards that particular distinction and won't appeal to it.

⁵¹ A somewhat similar line of argument is explored in Cokelet (2015).

can at one-time notice. Our moral systems should provide some allowance for this: for people whose particular perspectives, commitments, and values are slightly different from our own but which we can nonetheless appreciate.

We'll be looking at several cases here. For some readers, it will seem intuitive that the cases provide immediate support for the Character Condition. But, the argument is not that the cases are intuitive. Instead, the argument is that upon careful consideration, we cannot really capture what we value about the acts described without capturing how the person's character has informed and influenced the act. We'll begin with cases in which a person's moral character influences the rightness or wrongness of an act via their most deeply-held personal moral commitments.

Joyce the Lawyer: Joyce has recently passed the bar exam, and is considering her next career move. For years, she has volunteered at her state's Innocence Project, which seeks to exonerate people convicted of crimes they did not actually commit. She is deeply committed to helping people wrongfully convicted of crimes, but she knows that these cases typically take close to a decade from start to finish, and often fail even when there is overwhelming evidence of innocence. If she became a public defender, Joyce (correctly) believes that her work would impact many more people to a much greater degree, even when including the indirect impacts of the Innocence Project in terms of raising awareness of issues in our criminal justice system. Joyce is offered a job as a public defender, and as an attorney for the Innocence Project in her state. She decides to take the Innocence Project job.

We can begin by asking the question, even though Joyce's decision to take the Innocence Project job would not have as good consequences as taking the public defender job (which Joyce knew), and even though in both cases she would often be helping similar people from similar challenges, was it morally wrong for Joyce to take the Innocence Project job? I suspect that many readers (though, of course, not all) would respond that it was *not* wrong for Joyce to take the Innocence project job, despite

the fact that doing so would not produce the best consequences. Let's consider another similar case:

Sheila the Shelter Manager: Sheila is the manager at a non-profit, no-kill animal shelter specializing in rehabilitating abused animals. She started volunteering there 15 years ago because of her love and concern for the well-being of animals and her commitment to protect them. She gradually took on more and more responsibility, but that commitment never faded and in fact grew over time. One night, while Sheila takes a break from paperwork to go check on the animals, lightning strikes the animal shelter and the building quickly catches fire. Because of Sheila's location in the building, she can either go to the animal area, and save them (15 in total) from the fire by letting them out of their cages and outside, or she can go back to her office, and pick up an envelope that contains a very large cash donation that had been made earlier that day. Sheila believes (correctly) that she cannot save both the animals and the cash, and also (correctly) that the cash donation would allow the shelter to save hundreds more animals in the long term than it otherwise could have. Sheila decides to save the animals, and the cash burns in the fire.

Now, from a purely consequentialist perspective, Sheila arguably did the wrong thing. However, I suspect many of us would hesitate very much to claim that she *actually* did the wrong thing, even if we granted that (if we fleshed out the details of the case) she could have had a greater overall positive impact on animal welfare if she had saved the large donation.

Let's look at one more example. Since 2009, over 100 Tibetans self-immolated to protest Chinese occupation and treatment of Tibet (Branigan 2012; *The Economist* 2013). The Dalai Lama, the spiritual leader of Tibet, was blamed by the Chinese government for inciting these self-immolations, and was asked whether or not he would condemn them. Clearly, this is a situation where he had to weigh several conflicting moral considerations. Should he remain politically neutral, and evade answering? Should he drive support for a free Tibet at home by praising the self-immolators? But would that cause more people to harm or kill themselves through self-immolation? Is

the form of protest even politically effective? How can he tiptoe around angering China and thus instigating further incursions and retributions in and on Tibet? How should he weigh the *religious* significance of his response (fire has an important symbolic role in Buddhism where “Wake up! The world is on fire!” is an important injunction).

In the end, the Dalai Lama chose not to condemn the acts of self-immolation, ostensibly because doing so would hurt the vulnerable parents and families left behind by their family member’s choice to self-immolate (Branigan 2012). Now, certainly there is more than a little political savviness to this response, but it’s worth noting here that according to Tibetan Buddhism, the Dalai Lama is the incarnation of Avalokiteshvara, the bodhisattva of compassion. As such, it seems perfectly plausible that compassion count as an especially important moral consideration for him, and it further seems plausible to me that this matters for evaluating the Dalai Lama’s response. Assuming for the sake of argument that it was genuine, that the Dalai Lama really did feel compassion for the mourning family members and really did worry that using his platform to condemn the deceased would cause more suffering to those who were suffering most viscerally, it seems to me that his response was right – given his special moral commitments – in a way that it would not be right for others who did not hold those same special commitments.

Now, we can start to develop the argument. The main line of argument here is straightforward: what these cases illustrate or, at least suggest, is that we cannot capture our complete moral evaluation of all acts, what we value or disvalue about them, and how we understand the act itself without appeal to the character of the person who performed it. We value when people act on their *deeply-held* commitments, values, and

perspectives (when those themselves are morally acceptable). It is one of the things we cherish. Our theory of rightness should capture this fact: that is, if we value acting with integrity, our theory of rightness should capture it. Of course, there should be some limits here. The claim is not that *anything* is morally justified so long as it is especially morally important to someone. The action in question must be somehow morally defensible on its own (where by “defensible,” I do not mean that an action is conclusively right but that there are some, perhaps defeasible and perhaps ultimately unconvincing reasons to consider it as right). This argument provides no moral “get-out-of-jail-free” card for people to do terrible things that happen to accord with their (terrible) values. Further, the action in question must pertain to a situation in which the deeply-held moral commitments or principles are highly relevant, as compassion is a relevant moral commitment when considering how best to respond to mourning parents.⁵²

There are further lines of argument available here. It seems like the possibility for people to have special personal moral commitments is worth cultivating. Any moral system that does not do so risks running into a parallel of Bernard Williams’s famous integrity objection to utilitarianism: that it morally requires of a person to, in some circumstances, go against the moral projects and attitudes they take most seriously and hold most dear, and serves

⁵² One might worry here that I am conflating evaluating acts and evaluating motives, which are two distinct forms of normative evaluation. To some extent, I am doing that: I am claiming that we should not adopt a theory of act evaluation that is never informed by the agents’ motivations. It follows trivially from the Consequences Condition that a theory of rightness should not always be concerned about agents’ motivations, but a key commitment of the Character Condition is that our theory of rightness should be broad enough to be informed by certain motivations of some acts (specifically, motivations that are tethered to moral character in certain ways).

...to alienate him in a real sense from his actions and the source of his action in his own convictions. It is to make him into a channel between the input of everyone's projects, including his own, and an output of optimistic decision; but this is to neglect the extent to which his projects and his decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified. It is thus, in the most literal sense, an attack on his integrity. (Smart and Williams 1973, 117; qtd. in Chappell 2015)

That is, any moral system with standards of rightness that cannot accommodate at least some personal moral projects is liable to alienate that person from morality and from themselves.⁵³ Adopting the Character Condition can alleviate this liability. A somewhat related line of argument might appeal to Susan Wolf's (1982) famous discussion of "moral saints" – people who, in every case, acted in perfectly utilitarian or Kantian ways. Wolf points out that, though we highly value and want our loved ones to be moral people, we would not necessarily want them to be utilitarian or Kantian moral saints. A well-designed moral theory, according to perspectival naturalism, is constrained by our considered goals (including our goals for ourselves): and if something like the argument Wolf provides holds, it does seem that while we *are* happy for people to have their own personal moral projects (and thus, commitments and perspectives), we are less happy for people to have certain kind of maximizing moral systems that would preclude even the possibility of morally imperfect but personally important moral projects, values,

⁵³ This is especially the case for any maximizing theory of rightness, such as classical utilitarianism, according to which the right action maximizes some value. Satisficing theories of rightness (Slote 1984; see also Jamieson and Elliot 2009), according to which any act that meets a certain minimum standard is right are less susceptible to this type of worry. However, there is a tentative version of it that might be applicable: wherever the minimum standard is set, there will always be boundary cases that just *barely* fail to meet it. If we think personal moral projects are valuable at all, there can always be some that are just barely below the minimum standard, wherever it is, that we might think should not be treated as wrong. This, of course, depends on the minimum standard being relatively high – if it is relatively low than this response should be less convincing.

commitments and perspectives. We might further note that giving space in our theories of right and wrong for people to develop and own their own moral commitments, values, and perspectives (within the constraints we've noted) has a special place according to perspectival naturalism: to do so is to participate, part and parcel, in the moral project. What we are doing *in general* is creating the moral systems that best capture our most deeply-held values, commitments, and perspectives, and thus giving space for people to do this on their own *aligns* our moral system in a deep and important way.

Now, beyond valuing certain acts, in part, because they are performed out of the deeply personally held commitments, values, and perspectives of the people who did them, there is another way in which the unique moral character of a person should influence our evaluations of the morality of their acts. That is that part of what we value and cherish about certain acts is the *particular* way in which a certain person does them. Let's consider two examples.

Softball Sarah and Sandy: Sarah is 10 years old, and enjoys playing softball. She loves playing pitch and catch with her father, who works from home. During summer days, when Sarah does not have school, she frequently asked her father to play pitch and catch with her. Her father *always* apologizes, and says he cannot – he is just too swamped with work. And always within 10 minutes, her father changes his mind, takes a break from his work, and goes play pitch and catch with Sarah. One day at the playground, Sarah gets into an argument with a classmate and softball teammate of hers, Sandy. Sandy is bragging about her own father, and how every time she asks him to play pitch and catch with her, he agrees immediately, without any hesitation. When Sarah says that her father also always agrees to play pitch and catch, but never right away, Sandy says “Well that seems very bad. My dad always does the right thing and plays pitch and catch right away.” Years later, Sarah reflects that in fact part of what she cherished about her father's actions wasn't just that he was always willing to play pitch and catch with her, but also that he always said “no” first and then changed his mind.

The idea here is that part of what we cherish about certain acts is not merely the act itself, but the particular (but deep-seated ways) that certain people go about performing the act.⁵⁴ There can be individual differences in how we care for others, and we can cherish some of these individual differences. Lumping together what we value about the act and what we value about the way it was performed occludes this fact. The second example is a well-known, but possibly apocryphal story about George Washington, Alexander Hamilton, and Gouverneur Morris (as recorded by Martin Van Buren). Washington was famous for his serious, aristocratic bearing and temperament of “dignified reserve” – a bearing and temperament that was often at odds with, but just as often deeply-admired by his (to put it charitably) typically less formal American compatriots:

When the Convention to form a Constitution was sitting in Philadelphia in 1787, of which General Washington was President, he had stated evenings to receive the calls of his friends. At an interview between Hamilton, the Morrises, and others, the former remarked that Washington was reserved and aristocratic even to his intimate friends, and allowed no one to be familiar with him. Gouverneur Morris said that was a mere fancy, and he could be as familiar with Washington as with any of his other friends. Hamilton replied, “If you will, at the next reception evening, gently slap him on the shoulder and say, ‘My dear General, how happy I am to see you look so well!’ a supper and wine shall be provided for you and a dozen of your friends.”

The challenge was accepted. On the evening appointed a large number attended, and at an early hour Gouverneur Morris entered, bowed, shook hands, laid his left hand on Washington’s shoulder, and said: “My dear General, I am very happy to see you look so well!” Washington withdrew his hand, stepped suddenly back, fixed his eye on Morris for several minutes with an angry frown,

⁵⁴ One might worry here that there are in fact two acts here: the initial refusal to play pitch and catch followed by the acceptance of the offer to play pitch and catch. However, it is important to keep in mind here a key idea underlying all this discussion: that our evaluations of an act’s rightness should be informed as much as possible by what we value about it. An evaluation of the second act without any reference to the first, in this scenario, is incomplete, since it could not provide an adequate explanation of what is cherished about the first.

until the latter retreated abashed and sought refuge in the crowd. The company looked on in silence. (Van Buren 1867, 106)

As Van Buren goes on to explain, “It is without doubt true, that in his intercourse with public men Washington observed an extraordinary degree of dignified reserve, and there is every reason to believe that this invariable habit was natural to him, and in no degree assumed for effect” (Van Buren 1867, 107). What I want to emphasize about this anecdote is that, for many situations and for most people, reacting as George Washington here did would be frowned upon, to say the least. But here, in the later telling of the story, Washington’s response was cherished and admired, in large part because of how it illustrated particular aspects of George Washington’s character that were themselves cherished and admired.

Now we are in a position to make a more general point. In these two cases, we simply would not want or expect everyone to act in these ways – we certainly do not think that people are obligated to be as reserved as George Washington or to always change our mind before agreeing to play pitch and catch with a loved one. They are, in some sense, particular ways of being, but we cherish and admire them and the acts that grow out from them. This should influence our determination of rightness and wrongness in relevant cases: acts that might not be right (or as right) are made right (or more right) because of the way they depend on the particular person who performed them. What is more, there are further reasons for perspectival naturalists to try to build in an appreciation for individual differences in our moral behavior into our theories of rightness: as we saw in Chapter 3, given our psychologies, there are no perspectives we could adopt that could capture all of what we morally value. To notice and appreciate

all of what we value *requires* shifting perspectives and appreciating individual differences in moral behavior can help us to do that.

4.3.4 Value and Absurdity

According to perspectival naturalism, the best moral systems for us are those that are *best designed* for us. Thus far, I have offered several arguments that moral systems that meet the Character Condition are (ceteris paribus) well-designed for us: most importantly that it manages some of our basic evaluative tendencies well and that it helps us capture what we value about certain acts well. In Chapter 3, we saw another way in which moral systems could be well-designed for us: if they have positive externalities (or avoid the negative externalities of their competitors). That a moral system has some positive externality – some nice, (perhaps) unintended benefit – typically does not provide a *strong* reason in its favor, but it does provide *a* reason in its favor, according to perspectival naturalism. This point does need to be emphasized at the outset: my goal is not to provide an additional *decisive* argument in favor of the Character Condition here, but instead to introduce a small but important reason in its favor: that moral systems that meet the Character Condition provide an extra line of resistance to an important kind of moral absurdity.

At times in our lives, we find that our ability to care, to find value and meaning is greatly diminished, often because we have suffered a great loss, or are experiencing depression or an existential crisis or a feeling of absurdity. Philosophers in the analytic tradition have tended not to be tremendously concerned with issues relating to meaningless and absurdity. Consider, for example, a story recounted by Hare about a

house-guest of his reading a Camus novel and concluding that “nothing matters,” in which Hare glibly defeats the specter of meaninglessness:

...when we say something matters or is important, what we are doing, in saying this, is to express our concern about that something . . . Having secured my friend’s agreement on this point, I then pointed out to him something that followed immediately from it. This is that when somebody says that something matters or does not matter, we want to know whose concern is being expressed or otherwise referred to. If the function of the expression ‘matters’ is to express concern, and if concern is always somebody’s concern, we can always ask, when it is said that something matters or does not matter, ‘Whose concern?’ (Hare 1972)

And thus, Hare argues, so long as anyone is concerned about anything, something matters (this story is recounted in Parfit 2013, 410). But, anti-realists and (thus) perspectival naturalists cannot so easily dismiss an attitude of meaninglessness. For perspectival naturalists, values and reasons only have normative force from within a normative perspective, and if there is a category of perspectives in which concern over values and reasons is greatly diminished or eliminated, then there is no easy remedy from within that perspective. Of course, there is a tension here: people (with the possible exception of those experiencing the most severe forms of depression) rarely experience a *complete* inability to value or feel concerned about things, and they typically know that they used to value things and that others still do. Thomas Nagel has captured this tension well:

We cannot live human lives without energy and attention, nor without making choices which show that we take some things more seriously than others. Yet we have always available a point of view outside the particular form of our lives, from which the seriousness appears gratuitous. These two inescapable viewpoints collide in us, and that is what makes life absurd (Nagel 1971, 718)

Now, as Nagel rightly points out, some common arguments that our lives are absurd or meaningless are not very good – for example, that from a cosmic perspective our actions have such infinitesimally small effects that they are meaningless (Nagel

responds by asking why anything needs to be meaningful on a cosmic scale to be meaningful (Nagel 1971, 717), and a perspectival naturalist might ask the justification for adopting a cosmic perspective rather than a local one). But, for a perspectival naturalist, there is a special challenge here, because according to perspectival naturalism (and any anti-realist theory) there *is* no objective meaning or value in the world. So, in a certain sense, a sense of meaninglessness correctly represents the way the world really is. And further, since a sense of meaninglessness often diminishes our capacity to fully appreciate reasons, the fact that typical reasons taken to support the claim that life is meaningless or absurd are not very good reasons is somewhat beside the point. As Nagel rightly points out, when we avoid a sense of meaninglessness or absurdity, it is not really because we've reasoned our way out of it, but rather because we do *not* reason:

We do not step outside our lives to a new vantage point from which we see what is really, objectively significant. We continue to take life largely for granted while seeing that all our decisions and certainties are possible only because there is a great deal we do not bother to rule out. (Nagel 1971, 723)

That is, when we are not experiencing feelings of meaninglessness or absurdity, it is not for good epistemic reasons but because we just happen to not be occupying the perspectives in which reasons that would push us towards those feelings. There is reason to believe that moral systems meeting the Character Condition can help provide a small but important defense from shifting into perspectives that facilitate feelings of absurdity and meaninglessness.

To see how, it will help to take a perhaps surprising detour into communist-era Soviet-bloc absurdist dissident literature, such as the plays "The Garden Party" and "The Memorandum" by Vaclav Havel and the novel *The Joke* by Milan Kundera, which have explored a fascinating and terribly important route to the absurd. In these stories,

we find people who are tremendously concerned about human affairs while totally unconcerned about particular human beings; we find incredibly powerful and horrifyingly faceless, identity-denying normative systems (it is no surprise that the rallying cry of the 1968 Prague Spring was “Socialism with a human face”). In Kundera’s *The Joke*, a successful college student and enthusiastic supporter of the Communist Party is kicked out of college and forced into work brigades and coal mines for years of his life because he wrote a note teasing a classmate with a bad joke ending with “Long live Trotsky!” Havel, who is darkly fascinated with the idea of people creating new languages for moral, economic, or political purposes which become increasingly detached from reality (think here of double-speak in Orwell’s *1984*), emphasizes a similar theme in “The Garden Party” and “The Memorandum.” In “The Garden Party,” parents send their teenage son to attend a nearby garden party for the government’s Ministry of Liquidation with the hopes of befriending an influential patron there. At the garden party, everyone speaks in an absurdly bureaucratic invented language: the teenager quickly learns the language and becomes so adept at it as to be hired on the spot to oversee liquidating the Ministry of Liquidation. When returning home after the party, the teenager has taken aboard speaking in the invented bureaucratic language so fully and completely that his parents literally can no longer recognize him. In “The Memorandum,” an organization introduces a new language “Ptydepe” to its workers, aimed at making work more efficient by eliminating all homophones and entrenched connotations of and associations between words. A secretary who can read the new language but lacks a permit to translate it is fired for translating a memo written in Ptydepe for another worker who cannot read it. Shortly

thereafter, Ptydepe is replaced with a new language, which is then abandoned in favor of the original native language. What these stories all have in common is that they illustrate and exemplify a certain type of absurdity: the absurdity of a bureaucracy obsessed with human behavior but careless about actual, particular people. This is, to varying extents, a feature of many bureaucracies across times and cultures, and there is an important point to be drawn from them about our normative theorizing: moral systems need to inculcate not only dispositions to care *about* people but to care *for* them. It is a strange fact that moral systems can alienate us from other people *for the sake of* other people: they can occlude the harm we do to those in front of us by enshrining various moral principles or convincing us we are working “for the greater good.” In alienating us from what we typically do and should care about - particular people – some moral systems risk functioning themselves as faceless moral bureaucracies. While aiming to moralize us, they dehumanize us. This is very closely related to, if not a version of some of the concerns we saw in the previous section about integrity (raised by Bernard Williams) and moral saints (raised by Susan Wolf) (Smart and Williams 1973; Wolf 1982). The Character Condition helps us reduce this alienation: it better attaches our normative systems to our values and better directs our moral focus to the particular people committing acts. In doing so, we can better avoid falling into the trap of adopting dehumanizing moral systems. It does so by creating space in our moral evaluations of particular people and their unique personal and deeply-held commitments. It directs our attention to people and their particularities and away from principles and their uniformities. Now, of course there are some caveats here. The Character Condition is not a magic pill: I am not claiming that adopting it

would immediately and always help us avoid the kind of absurdity in question. The claim is weaker than that: that the Character Condition will not solve the problem, but it will help. Further, this argument applies only to one important type of absurdity. I do not claim that anything, much less the Character Condition could provide a general antidote for feelings of meaninglessness and absurdity in general (this is, of course, especially the case for depression).

4.4 Objections to the Character Condition

In the previous section, I provided several arguments in support of the Character Condition. I don't claim to have provided a complete argument in its favor: more could be said. But we have seen arguments along several trajectories in favor of the Character Condition: we do tend to evaluate actions, in some cases, based partially on character (I argued in section 4.2.1), doing so is consistent with our values (I argued in section 4.2.2), and there are benefits to a moral system that ties together our action and character evaluations (I argued in sections 4.2.2, 4.2.3 and 4.2.4). We should now consider some potential objections.

4.4.1 The Person-Neutral Objection

The first immediate objection to the Character Condition we'll examine states that a key condition on an adequate theory of rightness is a Person-Neutral Condition.⁵⁵ According to this condition, an adequate theory of rightness should tell us what *any* person, regardless of their particular personality, identity, or moral character, should do

⁵⁵ We have to be careful with this terminology. Typically an "agent-neutral" condition is that a theory weighs the goodness or badness of consequences for each individual equally: think here of Bentham's dictum "everybody to count for one, nobody for more than one" (Mill 1979). To avoid confusion that that condition, I use the term "person-neutral condition" here.

in a given situation. The Person-Neutral Condition is straightforwardly incompatible with the Character Condition. But, the Person-Neutral Condition has several issues, and we should not accept it. As always, we should start by emphasizing the fact that there are no fully-mind independent objective facts of the matter about whether theories of rightness match the Person-Neutral Condition or not, so strongly-held antecedent intuitions in support of one or the other cannot provide the entirety of the justification for either. And so, we should ask (as always) whether a well-designed moral system would treat rightness as meeting the Person-Neutral Condition. There are reasons to think it would not. If one of the things we want from the concept of rightness is to, at least in some circumstances, be action-guiding, we should not accept the Person-Neutral Condition. It cannot be fully action guiding. It will be the case in certain circumstances that I ought to do some act because of some feature of my particular personality, identity, or character – this is what I’ve argued. But a theory of rightness committed to the Person-Neutral Condition cannot tell me what I should do in such circumstances (and it cannot deny in a non-question-begging way that some situations might exist).⁵⁶ But, the main issue here is that a theory of rightness constrained by the Person-Neutral Condition is less flexible than it should be. On the view I’ve been advocating, we can legitimately ask (and provide answers for) the questions: “What would be right for *me* to do in these circumstances?” and “What would be right for any person to do in these

⁵⁶Now, one might try to mitigate this worry by distinguishing, in some cases, between what I morally ought to do and what is right for me to do, where what I ought to do is not person-neutral but what is right to do is. But, this leads to problems. First, it leads to the conclusion that I ought to perform some acts that are not right, which is an odd outcome. Second, in the design paradigm, it is not clear how this distinction is more effective than simply treating what I morally ought to do as equivalent to what is right to do.

circumstances?” without collapsing them, whereas a theory of rightness constrained by the Person-Neutral Condition cannot treat these questions (and their potential answers) as legitimately distinct. Further, the discussion of the Consequences Condition suggests an error theory for the Person-Neutral Condition: because there are some acts that genuinely would be wrong for any person to commit, we erroneously extrapolate to the Person-Neutral Condition. This error might be understandably exacerbated if the conditions in which a person’s identity, personality, and character should influence our determination of the rightness of their actions are not particularly common or salient to us, which, on the view I’ve advocated, is the case. Thus, a less radical condition, according to which some but not all acts should be labeled as right or wrong regardless of the person who performs them is much more flexible, seems much more plausible, and is consistent with the Character Condition.

4.3.2 The Heuristic Objection

A second and distinct line of objection goes something like this: as a practical matter, it is often very useful to utilize information about a person’s character when evaluating the rightness or wrongness of their actions. This is because we lack access to the important information about a person’s inner mental life that actually should help determine the action’s rightness – we don’t know exactly what the person was thinking and feeling when she made the relevant decision to act in a certain way, but considering a person’s character helps us guess. Thus, thinking about a person’s character when making act evaluations is practically useful, but in reality, doing so is nothing more than using a heuristic for guessing the non-character features of an act, such as the agent’s intentions and motivations that actually should determine how to evaluate it. So, the

Character Condition mistakenly welds what should merely be a practical heuristic onto a theory of rightness. This “heuristic objection” is, I think, one of the more interesting objections to the Character Condition, but I do think there is a satisfying, non-ad hoc response: this objection starts out on the right track, but fails to proceed all the way to its own logical conclusions, and if it did, it would actually arrive at the Character Condition. After all, if we think that a person’s motivations sometimes matter for rightness, we might also become convinced that whether those motivations were transient or deeply-held might matter, and also then that the perspective the person inhabited at the time of their decision is relevant, and so we might also think that the person’s deeply-held principles and values might matter as well. Before we know it, we’ve pointed out enough features that might be relevant that our list of features starts to look a lot like moral character. In other words, for this objection to succeed, it needs to show that only some small proper subset of mental states, properties, and features of a person that we sometimes think determine in part the rightness or wrongness of their action should count, and do so in a principled, non-ad hoc fashion. But it is very difficult to see how one could do this. Once we’ve allowed that motivations sometimes matter for rightness, we should carefully examine how we think they matter, which ones we value or disvalue, abhor or cherish, and why. And, if the arguments I’ve given in support of the Character Condition hold, it seems that we cannot fully specify this in all cases without capturing features of a person’s character.

4.4.2 The Situationist Objection and the Responsibility Objection

A third line of immediate objection to my central thesis is motivated by the well-known situationist critique of virtue ethics, and we will need to spend some time here

saying something about it. According to the basic version of this objection, we should not accept the Character Condition as a condition on a theory of rightness because moral character simply does not exist: our behavior is influenced far too strongly by external situational factors than it is by internal factors such as our personality or character (Doris 1998, 2002; Harman 1999; Machery 2010; Merritt 2000).⁵⁷ Now, we'll need to be a bit careful here: since situationist critiques are aimed most directly at virtue theories, they are typically most focused on whether virtues, understood as types of personality traits exist, rather than character more generally. For convenience, I'll follow that idiom in much of the discussion that follows, but the general idea of the objection I'm interested in is this: the same empirical evidence that was levied against

⁵⁷ We should be careful to distinguish here between two broad forms of situationist critique: direct and indirect critiques (see Sreenivasan 2017; Robertson 2018). Direct situationist critiques claim that situationist findings in empirical social psychology show that our psychologies are too influenced by random situational factors for us to think that our personalities are consistent enough to ground anything like virtues or moral character, at least on traditional ways of understanding virtues and moral character. These are the most familiar form of situationist critique, and have been put forward by Harman (1999), Doris (1998, 2002), Merritt (2000) and Machery (2010). The indirect critique, which was introduced by Doris (see Doris 2002, 146) and is the focus of (Sreenivasan 2017) is more subtle, and, again in short, says that according to our empirical evidence, such a staggeringly low number of people will ever be able to acquire virtues that a system of virtue ethics that says we ought acquire characterological trait virtues is unjustified. We'll only be focusing on direct situationist critiques here. There is a third type of critique, which has been offered by situationists (specifically Doris 2002; Ciorria 2014 develops and responds to this line of critique), but is not necessarily situationist in nature, that has been targeted at aspirational trait virtue theories (e.g. Hursthouse 1999). According to this type of critique, treating virtues as aspirational (rather than actual psychological dispositions) leads to problems of "excessive theoretical mediation" – in some situations for example, thinking "I ought to be empathetic towards that suffering person" is one thought too many. One might attempt to apply this critique to the Character Condition, but it is not exactly clear how. We saw in section 4.2.2 that as a matter of our moral psychology, we naturally and automatically to take a person's character into account when evaluating moral actions. And, because we're also committed to the Consequences Condition, it follows that in some situations we *shouldn't* engage in serious moral deliberation because the outcome is so obvious.

the existence of virtues could be levied against the existence of moral character, and if moral character does not exist, then we need to re-examine whether it should be appealed to in a well-designed theory of rightness.

There are at least three main strands of situationist critique, and they can each be applied with varying degrees of strength. They typically target “character trait” virtue theories on which virtues are personality traits that are stable, consistent, integrated, and situation-independent (though stability is often not treated as controversial). Stable character traits are “reliably manifested in trait-relevant behaviors over iterated trials of similar trait-relevant conditions” (Doris 2002, p. 22). For example, if sometimes I remain calm at the sight of snakes, and other times I panic, my courage qua snakes is unstable. Consistent character traits are “reliably manifested in trait-relevant behavior across a diversity of trait-relevant eliciting conditions that may vary widely in their conduciveness to the manifestation of the trait in question” (ibid.). If I behave courageously anytime I see snakes, but fail to behave courageously when my job is jeopardy, my courage is inconsistent. Doris’s (1998, 2002) original formulations of the situationist critique focused largely on consistency – he argued that while empirical evidence did not undermine the claim that our personality traits are stable in similar situations, it did undermine the claim that our personality traits are consistent across a variety of relevant situations. Maria Merritt (2000) and Edouard Machery (2010) have developed a version of the situationist critique targeting virtue theories that treat virtues as integrated (or unified) personality traits (see also Merritt, Doris, and Harman 2010). Machery argues that for virtue ethics to deliver on its promise to provide a robust account of moral character, it must posit that we “have some specific values, norms,

(first-and second-order) desires, beliefs, moods, emotion, etc” (Machery 2010, 228), and that they are integrated: “they are influenced by a common cause or they causally influence one another” (Machery 2010, 229). Merritt (2000) has done the most work towards developing the line of direct situationist critique that targets situation-independent character trait theories. She argues against character trait virtue theories that assume “motivational self-sufficiency of character” – the degree to which “the motivational structure of virtue [is], in maturity and under normal circumstances, independent of factors outside oneself, such as particular social relationships and settings” (Merritt 2000, 374) (p. 374).

And so, the worry is this: one of the key commitments of perspectival naturalism was that it was empirically adequate. But, if the notion of moral character we are appealing to in the Character Condition treats our character traits as consistent across wide-varieties of situations, our moral character as “motivationally self-sufficient,” or our moral psychologies as highly-integrated, we might fail to meet the empirical adequacy condition. We should note immediately that this likely forecloses one obvious type of response on behalf of the Character Condition: that character might be a useful fiction, and if it is, it might be beside the point whether it actually exists or not. Perspectival naturalists are extremely wary about useful fiction or noble lie arguments such as this: they clash both with our commitment to empirical adequacy and to our value of honesty, and so a well-designed moral system should avoid them as much as possible. But, there are better replies here available to perspectival naturalists supporting the Character Condition. Two of them follow extant objections and responses to situationist critiques. First, we *can* develop empirically-informed and

supported notions of character, such as Snow's (2009) theory of virtue built on CAPS models of personality (Mischel 1999; Mischel and Shoda 1995; see Russell 2009), Fleeson's Whole Trait Theory (Jayawickreme and Fleeson 2017; Fleeson and Jayawickreme 2015; Fleeson 2001), and Morton's (2013) "bounded" virtues. These accounts accept that situations influence our behavior but develop theories of personality traits and thus virtues that show how stability, consistency, integration, or situation-independence still exist in some more or less qualified fashion, depending on the view. A second type of response (which has often been paired with the first, denies that situationists critiques have identified a genuine problem for virtue ethics. The most common form of this response points out that virtue is supposed to be rare, and thus it is utterly unsurprising that experiments show so few people acting virtuously (Kamtekar 2004; Athanassoulis 2000). Similarly, Sreenivasan (2013) argued that the particular types of experiments situationists relied on focused so much on acts in particular instances that they failed to provide the type of empirical evidence needed to assess people's long-term character dispositions. Edward Slingerland (2011) has developed this type of response on behalf of the early Confucians. Situationists contend (and many virtue ethicists would grant) that "local" dispositional (i.e. stable but not consistent) traits cannot stand in for virtues: they are highly influenced by situational contexts, and they are not consistent or unified in any substantive sense (Slingerland 2011, 395; Doris 2002, 64, 1998, 507–8). First, Slingerland argues that empirical research actually suggests that some "global" (i.e. stable and consistent) personality traits (especially the "Big Five": openness, conscientiousness, extraversion, agreeableness, and neuroticism) do have important causal efficacy, and in many cases correlate to behavior more highly

than situational factors (Slingerland 2011, 397–98). Second, the distinction between local and global traits itself is not as clear-cut as the traditional situationist critique makes it out to be – at the very *best* local and global traits exist at different spots on the same continuum (Slingerland 2011, 399). Third, despite the haziness of the distinction between local and global traits, early Confucian virtues may actually be closer to local traits than global traits anyway (Slingerland 2011, 401–2)). Finally, in response to the worry that virtue ethicists are committed to a much higher correlation (1.0 or nearly 1.0) between character traits and behavior than empirical evidence suggests exists, Slingerland admits that such a high correlation is almost certainly impossible to attain, but he also responds that Confucian long-term, multifaceted moral cultivation allows one’s character traits to become more efficacious, while Confucian situational manipulation through the use of socially-engrained rituals makes it easier for people to act on their character traits. Confucian moral cultivation and social engineering simultaneously allow one to “lower the bar” and “jump higher” (Slingerland 2011, 408–10). But, going forward, the strongest response along these lines will likely be informed by changes in social and personality psychology driven by the replication crisis: much of the best evidence about the strength of situations on our behavior has failed to replicate either at all, or at least at anywhere near the effect sizes originally found (“Estimating the Reproducibility of Psychological Science” 2015). However, for us, the most important response to a situationist critique of the Character Condition is fairly simple: that the cases in which the Character Condition applies do not rely on any *extremely* robust notion of character according to which our moral characters are

extremely motivationally self-sufficient or integrated but on a more plausible empirically-informed notion of character.

A fourth, and somewhat related line of objection to the Character Condition grows out of what Galen Strawson (1994) has called “the basic argument”:

1. To be morally responsible for our actions, we must be morally responsible for our character, from which these actions stem.
2. We are not morally responsible for our character.
3. Therefore, we are not morally responsible for our actions.

Put aside, for a brief moment, concerns about improperly conflating evaluations of rightness and of moral responsibility and rightness. If the basic argument is sound, it might thus suggest potential trouble for the Character Condition: if we aren’t morally responsible for our character, should it really play any role in our evaluations of right or wrong? Of course, since according to the basic argument, we aren’t responsible for *any* of our acts, this general line of thinking should also suggest that we should never evaluate *any* act as right or wrong, and thus *any* condition on adequate theories of rightness or wrongness is fundamentally misguided. Instead, perhaps we should focus on a more targeted challenge for the Character Condition based loosely on the basic argument:

1. We are not morally responsible for our character.
2. The features of an action for which an agent is not morally responsible should not factor into evaluations of the action’s rightness or wrongness.
3. Therefore, the features of an action related to our character should not factor into evaluations of an action’s rightness or wrongness.

Now, an immediate response to this line of argument is to point out a fairly widespread belief that our evaluations of rightness and our evaluations of moral responsibility are simply two distinct categories of evaluation, and they needn’t influence each other (see, for example Arpaly 2003). The great majority of normative theories as well as common

sense morality takes some acts over which we do not have complete, direct control to be right or wrong: think here especially acts of negligence. There are further concerns with adopting this modified version of the basic argument, even if we grant the controversial claim that we are not at all responsible for our character. First, for a perspectival naturalist, moral responsibility itself is a constructed concept, and so for the argument to go through we would need to show that we should accept the notion of moral responsibility utilized in the argument, and it is not clear why we should do so.⁵⁸ Second, even if we are not responsible for our *entire* character, we are still be responsible (indirectly) for aspects of it. In the examples of the Character Condition I gave, the aspects of moral character that are relevant are aspects of character that *are* cultivatable, and in several of the cases (Washington and the Dalai Lama) actually were intentionally cultivated. While it is certainly true that we lack direct control over the influences on our moral character in early childhood, as our critical, evaluative, and self-reflective capacities develop we can and often do reflect on who we are as moral people and embark on projects of changing and developing different aspects of our personality and character: perhaps to be more diligent at certain tasks, or to be more empathetic to strangers, or so on. We can and sometimes do adopt radically different perspectives and moral commitments from those we were surrounded by when children.⁵⁹ This point aside, it is not clear that we *shouldn't* be morally interested in behavior over which we do not have direct control. We saw evidence from empirical moral psychology that a key reason we judge some spontaneous, impulsive acts (i.e.

⁵⁸ For example, Manuel Vargas has developed a theory of moral responsibility that treats moral character and agency as centrally important (Vargas 2013).

⁵⁹ Thanks to Nancy Snow for suggesting this point.

acts over which we do not have direct control) more strongly than others is what they reveal about our moral character because our moral character is predictive of our future moral behavior.

4.4.3 *Reasons and Excuses*

Let's briefly consider two final objections.⁶⁰ These objections are both based on the concern that the character condition cannot accommodate two very familiar types of situations: situations in which a person does the right thing for the wrong reasons, and situations in which a person excusably does the wrong thing. Let's look at these one by one.

First, according to both common-sense and many philosophical theories of rightness, it is possible to do the right thing for the wrong reasons or motivations. This will be most straightforwardly the case in familiar forms of consequentialism such as classical utilitarianism. Consider the following scenario: Misanthropic Mike mistakenly believes that if he pushes a blue button, it will cause every human being alive to be tortured for a thousand years. However, in fact, hitting the blue button cures cancer. When Misanthropic Mike hits the blue button, (many of us would be happy to say) he did the right thing but for the wrong reasons. Importantly, the claim that Mike performed the right act is perfectly consistent with the Consequences Condition, so if this claim is *inconsistent* with the Character Condition, I am doubly in trouble, since I claimed that the two conditions were consistent with each other. However, the Character Condition *is able* to accommodate the claim that it is possible to do the right thing for the wrong reason. This is for two key reasons: first, the Character Condition

⁶⁰ Thanks to Martin Montminy for raising these.

explicitly does not apply to every act, and second, in the acts for which it applies, it does not claim that moral character *alone* determines the rightness or wrongness of the action. Thus, the Character Condition is perfectly compatible with there being acts whose rightness has no relation to or influence from the moral character (or character-grounded motivations) of the agent who committed them. This helps us see how we can reply to the second objection here: that intuitively, some acts are morally wrong but excusable (for example, if the agent was non-culpably misinformed or was completely unable to do otherwise). For example, imagine Mistaken Mike, who hits the red button which tortures every human alive for a thousand years because he was misinformed by an otherwise trustworthy source that hitting the red button would cure cancer. But, one might worry, if the rightness of an action is grounded entirely in moral character and cannot be influenced by the exigencies of circumstance, there is no room for moral excuse. But again, this objection is based on a mistaken interpretation of the Character Condition. The Character Condition *does not* claim that the rightness or wrongness of an act (or other moral features of the act, such as whether it is excusable) is determined only by the moral character of the agent: in the cases in which moral character *should* influence our judgment of the rightness of an act, moral character is never claimed to be the *only* moral consideration for determining the rightness of the act.

4.5 Summary

In this chapter, I developed the argument for the Consequences and Character Conditions. The Consequences Condition states that the consequences alone of some, but not all acts should conclusively determine their rightness or wrongness. I argued that the Consequences Condition captures a great deal of our moral intuitions about the

importance of people's welfare without being radically reductionist and builds on natural evaluative tendencies to easily notice harm both immediately and to easily appeal to welfare-based considerations in moral deliberation and discussion. Further, the Consequences Condition helps utilize one of the main useful features of the concepts of right and wrong by providing a tool for quick, emphatic communication about the desirability or undesirability of certain courses of options when there is great moral urgency. I argued that the Character Condition also builds in natural evaluative tendencies that meld our evaluations of people and the acts they perform, that it helps us capture features of acts that we value and cherish most, and that it has a positive externality of better adding meaning and value into the world.

Chapter 5: Conclusion

5.1 In Summary

What makes a right act right or a wrong act wrong? In this dissertation, I have developed and defended a very partial answer to this question. I have argued that a minimally adequate account of rightness should meet *both* a Consequences Condition and a Character Condition. The combination of these two conditions is by no means intended to provide a complete account of rightness: other conditions apply. Yet, any account of rightness that does not capture the combination of these two conditions is, I have argued, flawed. The Consequences Condition states that for at least some (but not all) cases, the rightness or wrongness of an act should be determined *conclusively* by the act's consequences (i.e. how well the act promotes well-being). That is, for some actions, the consequences alone determine the action's rightness or wrongness. The Character Condition states that in some (but not all) cases, the moral character of a person should influence the determination of the rightness or wrongness of that person's actions. Like the Consequences Condition, the Character condition does not apply to every action. However, unlike the Consequences Condition, the Character Condition does *not* treat the rightness or wrongness of an action as ever fully or conclusively determined by a person's moral character: character is *a* consideration for rightness but never the only one. But, before I could begin the task of trying to justify these conditions on a theory of rightness, I first had to justify even beginning to undertake such a task when so much has already been said about theories of rightness and commitments to and intuitions about favorite theories run deep. To accomplish this first step, I had to take a detour through metaethics and the foundations of morality.

The study of morality is and must be the study of humanity. This is an old idea. For Aristotle, it meant that the goal of all intentional action, seeking the good, was for humans ultimately aimed at our conspecific *eudaimonia*. For the early Confucians, the connection was quite explicit: their most fundamental moral virtue was *ren* (仁), translated sometimes as “benevolence,” but more interestingly as “humaneness” or “humanity” (and though one should always be careful about inferring anything of philosophical importance from the composition of Chinese characters, it is worth pointing out that the symbol for *ren* is composed of the characters “人” for “person” and “二” for “two”). More recently, certain types of constructivist metaethical views have taken as a foundation for morality the normative perspectives we have evolved and constructed (Street 2010) and pragmatic naturalist theories (Kitcher 2011, 2012; Flanagan, Sarkissian, and Wong 2016) have revived Dewey’s claim that “moral conceptions and processes grow naturally out of the very conditions of human life” (Dewey 1998, 2:354). Of course, not every moral philosopher has agreed with this general sentiment. Plato, at least for a period, thought that morality consisted in acquiring and acting on knowledge of the forms of the good, which existed outside of and did not depend on human psychology. For Kant, the demands of morality are the demands placed on all (human or non-human) rational creatures, in Kant’s specific sense of “rational.” For classical utilitarians, the rightness or wrongness of an action depends not (directly at least) on our specifically human moral psychologies but on the amount of total pleasure or pain that action causes. For metaethical intuitionists such as Sidgwick or Moore, certain basic moral propositions are self-evidently true.

I presented an argument (in Chapter 3) that morality, if it is to avoid radical nihilism, *must* be substantially dependent on our conspecific *human* moral psychologies and perspectives, our values, abilities, and evaluative tendencies. It cannot be *strongly* realist (where strong moral realism is taken to be the claim that there are moral facts completely independent of our values and perspectives). That argument was a modification of recent “evolutionary debunking” arguments against moral realism (Street 2006; Vavova 2015; Joyce 2007). This “radical debunking argument” goes as follows:

1. **(Possibility)** At the beginning of moral inquiry, the space of epistemically possible objective moral facts is exceedingly vast.
2. **(Evolutionary Influence)** Our moral minds are influenced, guided, shaped, and constrained by our evolutionary history.
3. **(Empirical Principle)** The mere fact that a (putatively) concrete object, property, or relation is referenced in a conceptual truth does not imply that the named object, property or relation exists concretely in our world
4. **(Epistemic Hook)** To make progress past the beginning of objective moral inquiry, we need some epistemic hook into objective moral reality. Three main options for epistemic hooks are (a) fundamental moral conceptual truths, (b) a special moral epistemic capacity, or (c) epistemic tools and technologies to help narrow in on objective moral reality. However
 - a. **(Conceptual Truth)** Given (3) the Reference Principle, we should doubt that fundamental moral conceptual truths can serve as an epistemic hook.
 - b. **(Capacity)** Given (2) Evolutionary Influence, we should doubt that any capacity of moral perception we have would accurately track objective moral reality
 - c. **(Narrowing)** Given the previous arguments (4a and 4b), we should doubt that the epistemic tools and technologies we use in other domains of knowledge could be helpful in narrowing in on objective moral reality.
5. Therefore, if there are objective moral facts, we should doubt that we will ever discover them. In other words, we should believe in moral anti-realism.

The radical debunking argument begins by giving (strong) moral realism the benefit of the doubt and treating inquiry into the moral domain as similar to inquiries in other domains: at the beginning of any of them, there are a great many things that *could be*

true given what we know. For example, at the beginning of inquiry into meteorology it was epistemically possible that thunderstorms were created by the gods and at the beginning of inquiry into astronomy it was epistemically possible that the stars revolved around the earth. What evolutionary debunking arguments (whether traditional or radical) try to do next is show that, if we were relying on completely mind-independent moral facts, inquiry into the moral domain could not progress because it has key differences from (and different tools available than) other domains of inquiry. In traditional debunking arguments, this is because our basic evaluative tendencies and our moral reasoning are so heavily influenced by evolutionary factors which we know evolved because of the fitness advantages they provided relevant ancestors and not because they are truth-tracking. The radical debunking argument adds to this a general philosophical principle: that the mere fact that an object, property, or relation is named in a conceptual truth does not, on its own, imply that the object, property, or relation exists concretely in our world. For example, while it is a conceptual truth that unicorns have horns, nothing about that conceptual truth implies that unicorns exist concretely in our world. Applied to the moral or metaethical domain, this means that a whole branch of responses to traditional evolutionary debunking arguments of moral realism, that we can discover moral facts by applying reasoning to certain self-evident moral truths, is problematized. The argument utilizes this principle and concludes by examining several of the most promising avenues for inquiry into the fully-objective moral domain and contends that there is no good reason to think that epistemic progress could be made in any of them.

However, in denying the plausibility of strong versions of moral realism, neither traditional nor the radical debunking argument implies the truth of a strong form of moral nihilism. After all, it is a key facet of debunking arguments that we have our own values, evaluative tendencies, and normative standpoints. These are the foundations (descriptively *and* normatively) of moral systems. And thus, the radical debunking argument, in one sense, leads to a rather non-radical implication: that what we have actually been doing when we engage in genuine moral deliberation, debate, and discussion is, for the most part, what we should be doing since this *just is* the process whereby moral systems are developed and refined, based on our values and perspectives.

This, on its own, is not a novel idea. Several philosophers (e.g. Street 2010; Joyce 2007; Kitcher 2011, 2012; Copp 2007; Mackie 1991) have developed anti-realist and constructivist normative theories based on the belief that there are no discoverable fully mind-independent moral facts. Given evolutionary debunking arguments, one of the most promising categories of such views is Humean constructivism, according to which moral systems derive their justification ultimately from features of our practical or normative perspectives but, contra Kantian versions of metaethical constructivism, no substantive set of moral claims is derivable from *all* practical or normative perspectives (see Street 2010; Arruda 2017). I introduced and developed a version of Humean constructivism called “perspectival naturalism.” The main commitments of perspectival naturalism are:

1. **Naturalism / Descriptive Adequacy:** moral theorizing should be informed by our best descriptive understanding of the world; this understanding will be naturalistic,

2. **Anti-realism:** there are no *knowable* objective, fully mind-independent moral truths,
3. **Non-nihilism / Constructivism:** anti-realism does not entail the abandonment of morality; moral systems are ultimately justified not by being objectively true but via our moral standpoints,
4. **Perspectivalism:** an adequate model of the mind must treat the mind as perspectival; perspectives do not reduce without loss to sets of beliefs, and
5. **The Design Paradigm:** The best moral systems for us are the moral systems that are best designed for us.

I developed and made the case for perspectival naturalism across Chapters Two and Three. Then, in Chapter 4, I put perspectival naturalism to use to make the case for the Consequences and the Character Conditions:

The Consequences Condition: for at least some (but not all) cases, the rightness or wrongness of an act should be determined conclusively by the act's consequences (i.e. how well the act promotes well-being): though there may be competing moral considerations, no competing moral considerations could outweigh the consequences in determining the rightness or wrongness of the act.

The Character Condition: in some (but not all) cases, the moral character of a person should influence the determination of the rightness or wrongness of that person's acts.

Since so few normative theories adequately capture both of these conditions the combination of them is particularly interesting. Utilizing the account of justification provided by perspectival naturalism, I argued in Chapter 4 that theories of rightness that capture the Consequences and Character Conditions are *better designed for us* than theories of rightness that do not.

I began with the Consequences Condition and compared it to the *Never-Consequences Condition*, according to which the consequences of an action should *never* conclusively determine our judgment of the rightness of that action. I provided several arguments that the Never-Consequences Condition was worse designed than the more moderate Consequences Condition: that the Never-Consequences Condition failed to capture intuitions about certain “obvious choice” cases (in which the only difference

between two options was a massive difference in the harm done between them), that it failed to fulfill one of the most useful functions of the concept of rightness (serving as a clear and quick notice that a particular action is *especially* helpful or harmful), and finally that given empirical evidence that our moral minds are tuned to notice harm and to connect it with moral judgments, a theory of rightness should, lacking strong countervailing reasons not to, try to harness and utilize this tendency rather than try to override it.

This leaves open the possibility that we should prefer the *Always-Consequences* Condition (according to which the consequences of an act should always conclusively determine the rightness of an action) to the *Consequences* Condition. I began my case against the *Always-Consequences* Condition by providing an error theory for it: I pointed out that the *Consequences* Condition is able to capture all of the intuitive cases that might support the *Always-Consequences* Condition. Then, since the *Always-Consequences* Condition is incompatible with the *Character* Condition, I developed the case in favor of the *Character* Condition. I argued that the *Character* Condition nicely utilizes our natural evaluative tendencies, that it nicely captures what we value about certain acts of attachment and virtue (see Pettit 2015) and about certain acts influenced by individual differences in moral character, and finally that the *Character* Condition contributes to the project of avoiding a certain kind of moral absurdity.

5.2 Implications reconsidered

Now that we have developed an account of perspectival naturalism and the *Consequences* and *Character* Conditions in much more detail and provided the arguments in their favor, it will be helpful at this point to take a closer look than we did

in Chapter 1 at competitor metaethical and normative views. We'll begin by looking at competitor metaethical views to perspectival naturalism, and then we'll look at how the combination of the Consequences and Character Conditions sits with various important normative views.

While there is no perfect or universally agreed upon schema for categorizing metaethical views, there is fairly wide-spread consensus that they can be divided at the outset into realist and anti-realist categories (i.e. whether or not there are knowable, objective moral facts) and cognitivist and non-cognitivist categories (i.e. whether or not moral judgments and claims sometimes successfully report moral facts in the form of propositional beliefs or whether they express non-propositional attitudes broadly construed). While there is a great deal of overlap between these two, they are not co-extensive: realist theories tend to be cognitivist, but not all anti-realist theories are non-cognitivist. I'll begin by looking at several important realist, cognitivist metaethical views, then shift to non-cognitivist anti-realist views. Finally, I'll turn to the cognitivist anti-realist views and constructivist views that are most closely related to perspectival naturalism.

I'd like to begin by looking at non-naturalist and intuitionist views. These do quite often go together, but we can (for a brief moment) distinguish between them: non-naturalism should be taken to refer to the *metaphysical* claim that moral properties are not natural properties whereas intuitionism should be taken to refer to the *epistemological* claim that moral knowledge is self-evident, and thus not acquired like empirical knowledge (Ridge 2014). Of course, that distinction won't matter a great deal for us: the radical debunking argument spells problems for both non-naturalism and

intuitionism. Moral non-naturalism was defended most famously by G.E. Moore in *Principia Ethica* (1903) but has also been advocated by David Wiggins (1991, 1998) and John McDowell (1998, 1995). Intuitionism has received more attention recently, and versions of it have been defended by Shafer-Landau (2005), Michael Huemer (2005), Robert Audi (2001), and David Enoch (2011), to name a few (see Stratton-Lake 2014). It will not be necessary here to delve into each of these views in any great detail, since the point of departure between all of them and perspectival naturalism occurs at the same place. Consider Huemer's elaboration of ethical intuitionism (which takes as its first commitment what I've called "non-naturalism"):

Ethical Intuitionism holds that moral properties are objective and irreducible. Thus, 'good' refers to a property that some things (perhaps actions, states of affairs, and so on) have, independently of our attitudes towards those things, and one cannot say what this property is except using evaluative language ('good', 'desirable', 'should', 'valuable', and the like). Intuitionists also have an epistemological thesis, from which their doctrine gets its name: that at least some moral truths are known intuitively. The notion of 'intuition' is subject to interpretation, but at least this much is generally meant: some moral truths are known directly, rather than on the basis of other truths, but not by the five senses (we do not see moral value with our eyes, hear it with our ears, and so on). (Huemer 2005, 6)

We saw in Chapter 3 how perspectival naturalism will depart from non-naturalist and intuitionist views. Since the radical debunking argument is epistemological, rather than metaphysical, it has no qualms with non-naturalism per se: the disagreement occurs once we try to develop an account of how we could know non-natural moral facts. Both traditional and radical debunking arguments are aimed largely at this epistemic issue: if our moral intuitions have been shaped by evolution in ways that seem to track effective ways of improving fitness among our ancestors rather than truth, why should we trust at all that our moral intuitions are leading us towards mind-independent moral truth? But,

against this criticism, intuitionist views have a nice reply: if we have discovered certain self-evident moral truths, then we have no good reason to doubt them, since even if we have unreliable intuitions, self-evident is self-evident. Moral truths thus might be like truths of probability: we are not especially naturally good at discovering them, but a proof is a proof. The radical debunking argument goes beyond traditional debunking arguments here in more directly targeting metaethical intuitionism's commitment to *self-evident* moral truths. It does so, as I described in the last section, by casting doubt on the claim that the moral properties referred to in apparently self-evident moral truths exist in this world as described.

There have also been several naturalist forms of cognitivist realism. We'll focus especially here on non-reductionist naturalism, often called "Cornell realism" (see, for example Sturgeon 1986; D. Brink 1989; Boyd 1988), which claims that moral properties supervene on or are multiply realized by natural properties but cannot be reduced without loss to them (A. Miller 2013, 144). This view takes its cues from claims in scientific ontology that certain scientific kinds, such as *gene* or *catalyst* are not truly natural kinds, but they are still real insofar as they figure in our best explanations of the natural world (Boyd 1988). These kinds *do* supervene on natural kinds: there cannot be a change in something's genes without a corresponding change in its natural properties. According to the Cornell realists, moral properties work similarly:

Some kinds of natural properties are highly complex, and knowable only through the functional role they occupy. Consider, for instance, the property of being healthy. Being healthy isn't like being red; there's no way that healthy people look. Of course, there may be some characteristic visual signs of healthiness—rosy cheeks, a spring in one's step—but these visual signs are neither necessary nor sufficient for healthiness. These directly observable properties are only indications of healthiness. Healthiness is a complex natural property, wholly constituted by an organism's body being in the "proper"

configuration. Healthiness has a robust causal profile. There are many things that can cause or impede health by their presence or absence: food, water, disease, etc. And there are many things that will result from health in typical circumstances: energy, long life, etc...

The Cornell realists hold that goodness is exactly like healthiness in all of these ways (Boyd 1988). Like healthiness, goodness is a complex natural property that is not directly observable, but nonetheless has a robust causal profile. Like “healthiness”, “goodness” is not synonymous with any simpler set of more directly observable claims. Instead, “goodness” describes the functionally complex natural property that is the effect of certain characteristic causes, and the cause of certain characteristic effects. (Lutz and Lenman 2018)

There are some important differences between Cornell realism and perspectival naturalism. The first is subtle but important: the types of kinds and properties that Cornell realism treats as real (both moral and scientific) are, a perspectival naturalist will emphasize emphatically, social constructs. They are constructed out of nature, but they are constructed by us for us. They are real in the sense that all of our inventions are real, but they are not part of the apparatus of the natural world: they do not carve nature at its joints but carve it where we want it carved. The more important difference however is that the justification for these constructed moral properties, for a perspectival naturalist, does not come from their *explanatory* power at all.⁶¹ Instead, they derive their justification from how well the moral systems that contain them are *designed* for us.

The next set of metaethical views I’d like to examine briefly are anti-realist non-cognitivist views, according to which our moral claims and judgments do not report truth-apt beliefs about the world but instead express various non-cognitive mental states and attitudes that are not truth-apt. For example, according to Ayer’s emotivism,

⁶¹ Strictly speaking, perspectival naturalists should doubt that these properties, on their own, have any explanatory power at all: it is only people’s belief in or understanding of them that could have explanatory power.

If I say to someone, ‘You acted wrongly in stealing that money’, I am not stating anything more than if I had simply said ‘you stole that money’. In adding that this action is wrong, I am not making any further statement about it. I am simply evincing my moral disapproval of it. (Ayer 1952, 107)

And thus a sentence like “Stealing is wrong,” according to Ayer “has no factual meaning – that is, it expresses no proposition that could be true or false” (Ayer 1952, 107). Simon Blackburn’s quasi-realism aims to provide an account of how we could talk *as if* our moral claims and judgments were truth-apt and genuinely referred to real moral properties despite the fact that they do not (Blackburn 1993), while Allan Gibbard has developed a non-cognitivist view according to which moral judgments are expressions of an agent’s rational acceptance of certain norms that are themselves grounded in a non-cognitive account of rationality (Gibbard 1990). Perspectival naturalism differs from all these views in two important, related ways. First, perspectival naturalism is constructivist in a strong sense: it claims that we create moral systems to suit our needs, and so there is no necessarily fixed meaning to moral claims. Thus, it is perfectly possible that various moral claims might be best described (semantically) by any of these theories, and it is perfectly possible that various moral claims simply are not adequately described by any of them. Second, and more specifically, according to perspectival naturalism our moral claims and judgments can and should report moral truths and facts: these are, for a perspectival naturalist, truths and facts about what the best-designed moral systems for us would advise.

Finally, I should turn to anti-realist views. I won’t say a great deal here about them, since in Chapter 3 I already discussed them in greater detail than other metaethical views as well as some differences between them and perspectival naturalism. First, and most straightforwardly, cognitivist anti-realist views such as

Mackie's error theory (Mackie 1991) and fictionalist views such as those of Mark Kalderon (2005) and Richard Joyce (2001), according to which moral claims are truth-apt, never actually true, and are sometimes useful will have, for a perspectival naturalist, the same problem that non-cognitivist views have: there is no reason to think that, even if our *current* moral claims are systemically false that they always must be or are necessarily so. Moral claims made from constructivist moral systems could perfectly well be true, and so if we understood our own moral systems accurately, there would simply be no need to be fictionalist. Perspectival naturalism also differs in important ways from its closest relatives, constructivist anti-realist views, such as Kitcher's pragmatic naturalism. Perspectival naturalism differs from its Humean constructivist cousins because of its commitments to a fundamentally perspectival philosophy of mind and epistemology and to the Design Paradigm, according to which the best moral theories for us are the moral theories that are *best designed* for us. It differs from Kitcher's pragmatic naturalism (and any *functional* anti-realist theory, according to which moral systems should be evaluated by how well they fulfill some fairly narrow particular function) in that perspectival naturalism refuses to evaluate moral systems along a single functional trajectory: according to perspectival there is not, need not be, and should not be a *single* function of morality that *could* ground all moral evaluation.

I will now turn from metaethics to normative ethics and discuss in more detail than I did in Chapter 1 how the Consequences and Character Conditions are inconsistent with various major normative ethical views. I won't be claiming that the combination of these conditions is inconsistent with *all* forms of consequentialism, deontology, and virtue ethics – after all, there are a great variety of possible views in

each category. But I will try to make clear that the combination of the Consequences and Character Conditions is at odds – sometimes starkly, sometimes subtly – with many important variants of these theories.

Let's begin with consequentialism. Of course, there are so many versions of consequentialism in the literature that I cannot begin to address them all, so I will here focus on some of the most important categories of consequentialist views:

1. *Direct / Indirect*: Direct consequentialist theories, such as classical hedonistic utilitarianism, assess actions on how well they directly promote some value. Indirect consequentialist theories assess actions by how well the motive (e.g. R. M. Adams 1976) or the rule (see, for example Hooker 2015; Brandt 2002; Smart 1956) behind the action promotes some value.
2. *Actual / Counterfactual*: Actual consequentialist views assess actions solely on their actual consequences – not on their possible, or intended, or likely consequences (see Philip Pettit 2015; Railton 1984).
3. *Value / Well-being* - Hedonism, Preference Satisfaction, Perfectionism, Objective List: Consequentialist theories vary on what they take to be the fundamental or most important moral value. Hedonist theories such as classical utilitarianism focus on pleasure and pain, preference satisfaction theories (see Hare 1982; Singer 1975) focus on the satisfaction of actual or laundered subjective desires of agents, perfectionist theories focus on developing our human nature (see Hurka 1993), and objective list theories (see Nussbaum 2000) focus on identifying certain basic goods such as friendship, health, and education.
4. *Maximizing / Satisficing*: Maximizing consequentialist theories claim that the right action is the action that *maximizes* the relevant value (e.g. classical utilitarianism). Satisficing views claim this is too demanding, and instead claim that any action that meets some minimum threshold is the right action (e.g. Norcross 1997).
5. *Distribution of value*: Different consequentialist theories might aim at different distributions of their relevant values. To take two of the most well-known options, consequentialism might aim for the highest *total* amount of its relevant good, or it might aim for the highest *average* amount of its relevant good (see Driver 2012, ch. 3).

A great deal of other and finer distinctions between consequentialist views have been made (see D. O. Brink 2006; Driver 2012; Sinnott-Armstrong 2015). But, we have enough here to get started. First, I should note the Consequences Condition is, unsurprisingly compatible with a great many consequentialist views. It will disagree

largely only with consequentialist views that are unconcerned (or not fundamentally or essentially concerned) with the welfare of others, such as strong egoist views and some strong desire-satisfaction views. The more interesting disagreements are centered on the Character Condition. Now, chunks of the Character Condition are perfectly compatible with *some* consequentialist views. We already saw how Pettit (2015; 2018) developed a broadly consequentialist theory committed to the claim that the goodness of certain acts could not be fully specified without reference to a person's character. Further, it seems that certain scalar or satisficing consequentialist views will count as right the types of actions the Character Condition counted related to individual differences, since on my account, these actions typically did need to meet some minimal moral threshold. But, there are departures and disagreements. For example, the theory I've developed simply isn't a maximization theory. Some actions that are made right because of individual differences *would not* maximize total or average value in certain circumstances. Further, the Character Condition is incompatible with a fully direct and a fully actual consequentialism, since part of what we value or disvalue about a person's character is its propensity or probability to lead to good or bad consequences in the future and across certain counterfactual situations.

Like consequentialist views, there are a great variety of deontological views. Trivially, we might convert the Consequences and Character Conditions into moral rules or duties, and thus derive a deontological theory apparently compatible with them. But, there is a bit of a subtle worry here: part of what makes deontological theories distinct is that they take the right to be fundamental to and explanatory of the good (Rawls 1980). This might not be a *necessary condition* of a deontological theory, but it

is certainly a central commitment of many of them. However, the Consequences Condition explicitly denies this: it says for some cases, the *consequences* of an action (specifically, the consequences of an action on people's welfare) tells the entire story about whether the action is right or wrong. Other moral considerations might exist (and on my view, do exist), but at least some of the time, the consequences are all we need to conclusively determine the rightness of an action, and further, the consequences are what *makes* the action right or wrong. Deontological theories that do not allow this are inconsistent with the Consequences Condition. Of course, there are a great many more particular differences as well. To take an immediate example, the Consequences and Character Conditions will diverge from Kantian ethics in a variety of ways – most obviously in that they are interested in rightness and wrongness at all. But there are more differences. For example, (as we just saw) the rightness of the Consequences Condition does not derive in any way from rational duties, and it is unlikely that the individual differences aspect of the Character Condition could be justified on the universalizability formulation of the categorical imperative (since I could will without irrationality a world where it did not apply) or the humanity formulation (since it is not clear to me that ignoring individual differences amounts to treating another person merely as a means). Perspectival naturalism and the Consequences and Character Conditions *do* share some sensibilities with Ross's deontological view developed in *The Right and the Good* (Ross 1930), but there are important differences as well. The most obvious similarity, which I noted in early chapters, is that they both treat morality as a project of managing and weighing multiple individually defeasible and occasionally conflicting considerations against each other. For Ross, these are *prima facie* duties,

while for perspectival naturalism they are design considerations. There is a great deal of overlap between the relevant considerations on the two views, but they are far from identical. There are other important differences as well: in particular, Ross believes his prima facie duties derive their justification from being self-evident, but on my view, moral considerations derive their justification from being part of well-designed moral systems. Finally, let's briefly consider Scanlon's contractualism. According to this view, "An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement" (Scanlon 1998, 153). On this view, our moral obligations derive ultimately from our (counterfactual) rational and informed acceptance of a set of principles. Now, again, there will likely be a great deal of overlap in what the Consequences and Character conditions prescribe and what Scanlon's contractualism will prescribe, since it should be the case that informed people would rationally accept the best-designed moral systems for them – though it is possible these might come apart. However, there is an issue of priority here, and it illuminates an important difference between the views. On my view, *if* people were to rationally accept the Consequences and Character Conditions, it is *because* they are well-designed for us: that is from where their justification derives. There justification does *not* derive solely (and, potentially, does not derive at all) from whether we would agree to them in certain circumstances.

Let's now turn to virtue ethical views. It will helpful here to begin with a distinction introduced by Michael Slote (2001) between agent-prior and agent-based virtue ethical theories. Agent-prior virtue ethical theories include features of agents such

as their virtues or moral character as *part of* the foundation for overall moral theories while agent-based theories treat features of agents such as their virtues as *the entire* foundation for overall moral theories (see Slote 2001). Thus, Aristotle and Aristotelians typically espouse agent-prior views, since virtue-related considerations are part of the story for determining which actions are right, but *eudaimonia* is also fundamentally important insofar as it directs and constrains our theory of what virtues are. Slote, on the other hand, defends an agent-based view, for which the entire moral system is grounded in the virtuous motives of individuals – specifically an agent-based ethics of caring (Slote 2001). There is an obvious tension here between agent-based theories and the Consequences Condition – though Slote’s ethics-of-care based view will *prescribe* the same actions that the Consequences Condition would, the Consequences Condition is meant to be understood and supported entirely independently from agential considerations. The massive differences in welfare tell the entire story, and any reference to agential features is superfluous, according to the view.

There is considerably less friction between non-agent-based virtue ethical theories and the view that I have developed. The Consequences Condition is compatible with many of them, insofar as they allow that an act could be right solely because of its impact on welfare (and this is what would inform a virtuous person in making their virtuous decisions), they are in perfect agreement, on this point at least. There will be some important smaller differences worth discussing. First, let’s begin with Aristotelian and neo-Aristotelian *eudaemonist* virtue ethics (see especially Hursthouse 1999). Perspectival naturalists will doubt that happiness or flourishing can really get us everything we need for a well-designed moral system. After all, one of perspectival

naturalism's main criticisms of its anti-realist and Humean constructivist cousins is that they often ground normativity in a single or small set of *functions*. Flourishing as a ground for a normative theory has the important strengths of being naturalistic and, by definition, unobjectionably worth having. While I believe that happiness and flourishing *should* be absolutely central to a well-designed moral system, there are simply other moral considerations. To be more specific, my own worry here is not that Aristotelian virtue ethical views fail to capture these other moral considerations, but instead that they've snuck them in – that there are moral considerations that can't justifiably be grounded solely in conspecific *eudaimonia* (this worry will transfer to several recent naturalist virtue ethical views such as those of Phillipa Foot (2001) and Michael Thompson (2008) – when we are being careful about exactly how we are evaluating putatively species-related goodness, we will see that our myriad values are actually doing the work).

Finally, I'll briefly turn to Linda Zagzebski's exemplarist virtue theory (2010, 2017). This view grounds moral concepts (such as those of rightness) ultimately in moral exemplars and our admiration for them. I should point out immediately one feature of the view that will be particularly impressive for perspectival naturalists – the connection it draws between exemplars, admiration, and normativity is an extremely well-designed component of a moral system. Since we tend to want to emulate those we admire (i.e., exemplars), the view builds in normative force to its prescriptions in an especially ingenious way. Further, at least aspects of the Character Condition are very friendly towards exemplarism, since individual differences in moral exemplars is a key part of the richness of the view. There are some, again, subtle points of departure. First,

as we have seen repeatedly, insofar as the Consequences Condition is justified solely on welfare-related grounds, without any mention of agency or exemplars, the two views may differ in what they take to be adequate foundational moral reasons. More interestingly, it seems to me that there are certain design considerations for moral systems that are difficult for exemplarism to capture because they are unlikely to correlate with differences in admirability. Consider our earlier discussion of *xenia*, the ancient Greek value of strict obligations for good hospitality. I argued that given the circumstances in daily life in ancient Greece, this especially strict moral obligation towards hospitality is more important than it might be in other moral systems (obligations to be hospitable to outgroup members are *never* remotely bad – they are simply even more urgent in circumstances in which people are more likely to encounter lots of outgroup members). But, it isn't clear to me that exemplarism and admiration can or would agree with this design consideration – there is likely not any relevant difference among exemplars in their hospitability (one would hope that they are all quite hospitable). Further, if we claimed that in certain cultures, people *might* admire these types of special circumstantial values more, and thus exemplarism can capture them, there is a priority worry. The reason people admire them more (e.g. why a person in ancient Greece might be more prone to admire a person who was more hospitable) is *because* that value is especially useful and important in that circumstance – that moral systems that better promote that value are especially well-suited to those circumstances.

5.3 Directions for future research

I would like to finish by briefly suggesting some potential avenues for future research based on the groundwork that I have laid in this dissertation. Of course, a great

deal more could be said for, against, and about the arguments I gave for the Character and Consequences Conditions. In particular, I think that the version of ethical absurdity I described at the end of Chapter 4 deserves much more attention and could ground several fruitful conversations. However, I want to look farther afield here, and instead focus more on topics I did not discuss in *any* detail, including some more things that perspectival naturalists could say about a theory of rightness, and some more things perspectival naturalists could say in general.

I emphasized throughout this dissertation that the Character and Consequences Conditions were not meant to provide a complete account of rightness, but I did not suggest in any detail what other plausible conditions on an adequate theory of rightness might be. Some further conditions are likely entailed by the Character and Consequences Conditions. For example, a Motivations Condition, according to which the motivations of the agent committing the act should sometimes influence our determination of the act's rightness or wrongness is implied by elements of the Character Condition, which takes into account individual differences in character-driven motivations. But, it seems likely that a plausible perspectival naturalist case could be made for other types of motivations influencing some of our judgments of right and wrong. More interestingly, there also might be a plausible perspectival naturalist case for a deontological condition on our theories of rightness – that is, “deontological” in the somewhat loose (but useful) sense of an allowance for certain rules that are resistant to straightforward cost-benefit (i.e. consequentialist) style justification. Consider Robert Frank's (1998) account of “commitment devices.” Frank (1998) was interested in certain types of cases that present challenges for models of human behavior as rational

and self-interested. In these cases a person appears to act against her own self-interest, without clear ancillary gains: a lawyer skips out on \$600 of billable hours to fight a \$100 traffic ticket she thought was unfair, we leave tips in restaurants we will never visit again, and we sometimes commit acts of revenge that are likely to have devastating costs to ourselves and others we care strongly about (Frank 1998, 2–3). Frank argues that this behavior *is* irrational, but usefully so, because it can help solve “commitment” problems that are unsolvable by purely rational self-interested behavior. The idea is that we develop “commitment devices” – emotions and attitudes and reputations that publicly commit us to behaving in certain irrational ways, and because we are known to be willing to act in these irrational ways, they serve as a type of protection:

Consider a person who threatens to retaliate against anyone who harms him. For his threat to deter, others must believe he will carry it out. But if others know that the costs of retaliation are prohibitive, they will realize the threat is empty. Unless, of course, they believe they are dealing with someone who simply likes to retaliate. Such a person may strike back even when it is not in his material interests to do so. But if he is known in advance to have that preference, he is not likely to be tested by aggression in the first place. (Frank 1998, 5)

Now, it is important to note that Frank is not developing a specifically moral argument here – his discussion is framed in terms of rationality and interests. But, using the design paradigm, we can apply it. One special problem that we face as humans is the following kind of immoral act: taking advantage of another person to a degree just slightly less than it is rationally worth it for them to fight back against. Think here of police who are more likely to give tickets for traffic violations to people from other states (who are unlikely to drive back to dispute the ticket in traffic court), or cable companies who charge incorrect fees but develop a byzantine and purposefully frustrating customer service system such that, for most customers, it is better worth their

time to just pay the incorrect fees and be done with it. If Frank is correct, and there are a great many similar types of situations that require people to act on commitment devices in ways that definitely do not lead to their short-term overall benefit, and on occasion will not even lead to their long-term overall benefit, then there might be good design reasons for a theory of rightness to allow and even encourage, in some cases, adherence to certain moral principles *despite* their immediate and even long-term consequences (though of course, any such principles should be compatible with the Consequences Condition).

Stepping back from theories of rightness, perspectival naturalism provides new and interesting opportunities for developing, understanding, and evaluating moral theories in general. It would do so through two of its identifying features: its commitment to psychological perspectivalism in evaluating theories and its commitment to the design paradigm.

The perspectivalism of perspectival naturalism can be utilized to better understand ethical theories and the arguments proponents give in their support. To take a single example of both, consider Peter Singer's famous rehearsal of Peter Unger's "Bob's Bugatti" case:

Bob is close to retirement. He has invested most of his savings in a very rare and valuable old car, a Bugatti, which he has not been able to insure. The Bugatti is his pride and joy. In addition to the pleasure he gets from driving and caring for his car, Bob knows that its rising market value means that he will always be able to sell it and live comfortably after retirement. One day when Bob is out for a drive, he parks the Bugatti near the end of a railway siding and goes for a walk up the track. As he does so, he sees that a runaway train, with no one aboard, is running down the railway track. Looking farther down the track, he sees the small figure of a child very likely to be killed by the runaway train. He can't stop the train and the child is too far away to warn of the danger, but he can throw a switch that will divert the train down the siding where his Bugatti is parked. Then nobody will be killed—but the train will destroy his Bugatti. Thinking of his joy in owning the car and the

financial security it represents, Bob decides not to throw the switch. The child is killed. For many years to come, Bob enjoys owning his Bugatti and the financial security it represents. (Singer 1999; Unger 1996, 135–36)

Unger and Singer both point out that if one morally condemns Bob in this case, one should also morally condemn discussions not to donate large amounts of money to charities that would save a huge number of children's lives. Many philosophers are tempted to treat such arguments as a straightforward argument from analogy: Case A has moral properties X, Y, Z, Case B also has moral properties X, Y, Z, so Case A and B are similar in all morally relevant respects, so one's moral evaluation of Case A should match one's moral evaluation of Case B. There should be no doubt that this is part of what arguments like this try to show, but it is by no means all that arguments like this try to do. A great argument from analogy also aims to push us to shift our perspective: to see certain things in a new way. It does so by describing the case in certain ways, using semantically- and associatively-loaded terms like "market value," "financial security," Bugatti," "*small* figure of a child" [emphasis added], with connotations and patterns of association that percolate throughout our perspectives, that direct our attention and make certain features of the situation appear more salient and central. Importantly, all of this pushes us to understand moral theories in a different way as well: Singer and Unger and utilitarians in general are not only aiming to get us to *believe* that we should, for example, save the child by sacrificing the Bugatti, or donate lots of money to save children, but to see people failing to save children through charitable donations *as* Bob refusing to sacrifice his Bugatti. They are encouraging us to adopt a certain perspective. When we realize and genuinely appreciate this point, it

greatly enriches our understanding of ethical theories and it further opens hosts of new questions and opportunities for new and useful work in ethics and moral epistemology.

In particular, it opens new opportunities for evaluating and criticizing ethical theories insofar as they do not merely make moral claims but also express and direct moral perspectives – treating certain features of the world as more morally salient or central while occluding others, prioritizing certain associations and inferences while dampening others, encouraging us to see certain things as similar and others as different. We can appreciate, or find problematic, the way that different ethical theories direct our attention. For example, we might appreciate how classical utilitarianism shifted attention to the suffering masses, but we might also worry how it simultaneously shifts our attention away from individual suffering that benefits the “greater good.” In such cases, an ethical theory’s claims and the perspectives it enshrines will go together, but, importantly, sometimes they might be in tension. To take just one example, think here of Kant’s claims about the fundamental importance of the dignity of all rational persons combined with his perspective that saw only a small minority of humans *as* rational persons.

The perspectivalism in perspectival naturalism also opens up and helps justify an interesting type of philosophical criticism of certain ethical claims and positions: that they are non-accidentally tethered to or even dependent on problematic perspectives.

Think here of Elizabeth Anscombe’s argument against a certain brand of consequentialism:

But if someone really thinks, in advance, that it is open to question whether such an action as procuring the judicial execution of the innocent should be quite excluded from consideration – I do not want to argue with him; he shows a corrupt mind. (Anscombe 1958, 14)

This type of argument can make quite a bit of sense to a perspectival naturalist when interpreted perspectivally: different ethical perspectives treat different moral questions as open and closed, as central or peripheral, as vital or unimportant. The mere fact that a certain perspective sees *certain* moral questions as open might cause us to doubt that perspective or the claims derived therein. Of course, one might respond that from a *philosophical* perspective, *every* question should be open. This might be the case, but I think that it is certainly worth debating, and having a rich understanding and appreciation for how perspectives shape and direct our thought is absolutely necessary for finding an adequate resolution to that debate. It seems to me that there are likely other types of arguments in philosophy that might be criticized along these perspectival grounds, or that existing criticisms might be enriched by translating them into a perspectival mode. To take just one example of a discussion that is or has been ripe to be so enriched, consider “anti-theodicies” – arguments that responding to the problem of evil is itself systemically immoral (see Betenson 2016; Tilley 1991; Phillips 2005). These arguments are often explicitly framed in perspectival terms: “Theodicy adopts too detached a perspective” (Betenson 2016, 58) , “the ideal observer [God] is *inhuman* in the worst sense of the word” (Gleeson 2012, 13), “[theodicy] betokens an irremissible moral blindness” (Surin 1986, 50; Betenson 2016, 59–60). A better understanding and great appreciation for the importance of perspective in our psychologies and moral psychologies might strengthen these types of arguments. In particular, it seems that it might powerfully supplement objections to free will theodicies if, for example, one shows that a consistently-held perspective throughout the sacred texts informing the theodicy shares the same type of fundamental issue that the

free will defenses do (namely, a callous disregard for bystanders and victims), then there might be a special worry for such defenses. Importantly, more recent work in this domain is much less susceptible to this kind of worry in part by adopting perspectives that in which suffering and evil are much more visible and salient (see for example M. M. Adams 1999; Stump 2010).⁶²

Perspectival naturalism's Design Paradigm could also ground numerous contributions to current ethical and philosophical discussions. Most obviously, we might ask about moral concepts other than rightness, such as goodness and badness, evil, and various character-related concepts such as the virtues what the *best-designed* version of them would be for us. Perspectival naturalism and the Design Paradigm would be especially fruitful for developing virtue theories.

Finally, perspectival naturalism and the design paradigm could be fruitfully applied to issues in epistemic normativity. Importantly, just as in the case of morality, perspectival naturalists will emphasize that design considerations are not particularly new in normative epistemology. Let's consider two brief examples. In his *Novum Organum*, Francis Bacon developed a proto-account of cognitive biases and heuristics, which he called "idols" (*Novum* 1.XXIII-XLIV). For example, Idols of the Tribe are conspecific biases and Idols of the Den are biases derived from contingent features our particular upbringing and education (*Novum* 1.XLI-XLII). Importantly for Bacon, the sciences derive a large part of their justification from the fact that they are so well designed to deal with and overcome the problematic natural "idols" that hamper better understanding of the world. Karl Popper's famous argument for falsifiability as a key tenant of science (Popper 2002) can be fruitfully understood along similar lines: a key

⁶² Thanks to Linda Zagzebski for suggesting that I emphasize this point.

mistake that we tend to make in developing and supporting theories is to slip into unfalsifiability, and so a scientific method that avoided this at all cost is well-designed to fix an error that we are especially prone to. Perspectival naturalists will treat these types of arguments as instances of tremendous insight into how normative epistemology can and should be done: by better understanding who we are, what we want, and the best way to get there.

References

- Adams, Marilyn McCord. 1999. *Horrendous Evils and the Goodness of God*. Melbourne: Melbourne University Press.
- Adams, Robert Merrihew. 1976. "Motive Utilitarianism." *Journal of Philosophy* 73 (14): 467–81.
- Alfano, Mark. 2016. *Moral Psychology An Introduction*. Malden, MA: Polity Press.
- Alicke, Mark D. 2012. "Self-Injuries, Harmless Wrongdoing, and Morality." *Psychological Inquiry* 23 (2): 125–28.
- Alter, Adam. 2017. *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked*. New York: Penguin Press.
- Anscombe, G.E.M. 1958. "Modern Moral Philosophy." *Philosophy* 33 (124): 1–16.
- Aristotle. 2014. *Nicomachean Ethics*. Translated by C.D.C. Reeve. Indianapolis, IN: Hackett Publishing Company.
- Arpaly, Nomy. 2003. *Unprincipled Virtue*. Oxford: Oxford University Press.
- Arruda, Caroline T. 2017. "The Varieties of Moral Improvement, or Why Metaethical Constructivism Must Explain Moral Progress." *Ethical Theory and Moral Practice* 20: 17–38.
- Athanassoulis, Nafsika. 2000. "A Response to Harman: Virtue Ethics and Character Traits." *Proceedings of the Aristotelian Society* 100: 215–22.
- Audi, Robert. 2001. "A Kantian Intuitionism." *Mind* 439 (1): 601–35.
- Ayer, A.J. 1952. *Language, Truth, and Logic*. New York: Dover Publications.
- Beilby, James, ed. 2002. *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*. Ithica, NY: Cornell University Press.
- Betenson, Toby. 2016. "Anti-Theodicy." *Philosophy Compass* 11 (1): 56–65.
- Blackburn, Simon. 1993. *Essays in Quasi-Realism*. New York: Oxford University Press.
- Bommarito, Nic, and Alex King. Forthcoming. "Matilal's Metaethics." In *Neglected Voices at the Foundation of Ethics*, edited by Collin Marshall. Routledge.
- Boyd, Richard. 1988. "How to Be a Moral Realist." In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord. Ithica, NY: Cornell University Press.
- Brandt, Richard. 2002. "Towards a Credible Form of Utilitarianism." In *Consequentialism*, edited by Stephen Darwall, 207–35. Oxford: Blackwell.
- Branigan, Tania. 2012. "Dalai Lama Says He Must Remain Neutral on Self-Immolations." *The Guardian*, July 9, 2012, sec. World news. <http://www.theguardian.com/world/2012/jul/09/dalai-lama-neutral-on-self-immolations>.
- Brink, David. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Brink, David O. 2006. "Some Forms and Limits of Consequentialism." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 380–423. Oxford: Oxford University Press.
- Brosnan, Kevin. 2011. "Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?" *Biology & Philosophy* 26 (1): 51–64.
- Brown, Campbell. 2016. "The Rightest Theory of Degrees of Rightness." *Ethical Theory and Moral Practice* 19 (1): 21–29.

- Camp, Elisabeth. 2006. "Metaphor and That Certain 'Je Ne Sais Quoi.'" *Philosophical Studies* 129: 1–25.
- . 2013. "Slurring Perspectives." *Analytic Philosophy* 54 (3): 330–49.
- Campbell, Richard, and Victor Kumar. 2012. "Moral Reasoning on the Ground." *Ethics* 122 (2): 273–312.
- . 2013. "Pragmatic Naturalism and Moral Objectivity." *Analysis* 73 (1): 446–55.
- Chappell, Sophie Grace. 2015. "Bernard Williams." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2015. Metaphysics Research Lab, Stanford University.
<https://plato.stanford.edu/archives/spr2015/entries/williams-bernard/>.
- Choi, Incheol, and Richard E. Nisbett. 1998. "Situational Salience and Cultural Differences in the Correspondence Bias and Actor-Observer Bias." *Personality and Social Psychology Bulletin* 24: 949–60.
- Ciurria, Michelle. 2014. "Answering the Situationist Challenge: A Defense of Virtue Ethics as Preferable to Other Ethical Theories." *Dialogue* 53 (4): 651–70.
- Clarke-Doane, Justin. 2012. "Morality and Mathematics: The Evolutionary Challenge." *Ethics* 122 (2): 313–40.
- Cokelet, Bradford. 2015. "Dispositions, Character, and the Value of Acts." In *Character: New Directions from Philosophy, Psychology, and Theology*, edited by Christian B Miller, R. Michael Furr, Angela Knobel, and William Fleeson, 233–50. Oxford: Oxford University Press.
- Copp, David. 2007. *Morality in a Natural World*. Cambridge, UK: Cambridge University Press.
- Crisp, Roger. 2015. "A Third Method of Ethics?" *Philosophy and Phenomenological Research* 90 (2): 257–73.
- Critcher, Clayton R., Yoel Inbar, and David Pizarro. 2013. "How Quick Decisions Illuminate Moral Character." *Social Psychology and Personality Science* 4: 308–15. <https://doi.org/10.1177/1948550612457688>.
- Dancy, Jonathan. 2006. *Ethics Without Principles*. Oxford: Clarendon Press.
- Datson, Lorraine. 1992. "Objectivity and the Escape from Perspective." *Social Studies of Science* 22 (4): 597–618.
- Dewey, John. 1998. *The Essential Dewey: Ethics, Logic, Psychology*. Edited by Larry A. Hickman. Vol. 2. Indiana University Press.
- Doris, John. 1998. "Persons, Situations, and Virtue Ethics." *Nous* 32 (4): 504–30.
- . 2002. *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.
- Dougherty, Alicia. 1998. "Gluing the Web Together: An Interview with Larry Wall." *ZDNet*, April.
- Driver, Julia. 2012. *Consequentialism*. New York: Routledge.
- Elgin, Catherine. 2002. "Art in the Advancement of Understanding." *American Philosophical Quarterly* 39 (1): 1–12.
- . 2010. "Keeping Things in Perspective." *Philosophical Studies* 150 (3): 439–47.
- Enoch, David. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
- "Estimating the Reproducibility of Psychological Science." 2015. *Science* 349 (6251). <https://doi.org/10.1126/science.aac4716>.

- Filler, Martin. 2011. "When Modernism Entered the Kitchen." *The New York Review of Books*, January 21, 2011. <https://www.nybooks.com/daily/2011/01/21/when-modernism-entered-the-kitchen/>.
- Fiske, Susan T., Amy J.C. Cuddy, and Peter Glick. 2007. "Universal Dimensions of Social Cognition: Warmth and Competence." *Trends in Cognitive Sciences* 11 (2): 77–83.
- Fitzpatrick, William. 2008. "Morality and Evolutionary Biology." *Stanford Encyclopedia of Philosophy*.
- Flanagan, Owen, Hagop Sarkissian, and David Wong. 2016. "Naturalizing Ethics." In *The Blackwell Companion to Naturalism*, edited by Kelly James Clark, 16–32. John Wiley & Sons.
- Fleeson, William. 2001. "Toward a Structure- and Process-Integrated View of Personality: Traits as Density Distributions of States." *Journal of Personality and Social Psychology* 80 (6): 1011–27.
- Fleeson, William, and Eranda Jayawickreme. 2015. "Whole Trait Theory." *Journal of Research in Personality* 56: 82–92.
- Foot, Philippa. 2001. *Natural Goodness*. Oxford: Oxford University Press.
- Fraasen, Bas van. 2008. *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press.
- Frank, Robert. 1998. *Passions Within Reason: The Strategic Role of the Emotions*. W.W. Norton & Company.
- Gibbard, Alan. 1990. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
- Gleeson, Andrew. 2012. *A Frightening Love: Recasting the Problem of Evil*. Basingstoke: Palgrave Macmillan.
- Gray, Kurt, and Jonathan E. Keeney. 2015. "Disconfirming Moral Foundations Theory on Its Own Terms: Reply to Graham (2015)." *Social Psychology and Personality Science*, 1–4.
- Gray, Kurt, and Chelsea Schein. 2016. "Moralization and Harmification: The Dyadic Loop Explains How the Innocuous Becomes Harmful and Wrong." *Psychological Inquiry* 27: 62–65.
- Gray, Kurt, Chelsea Schein, and Adrian F. Ward. 2014. "The Myth of Harmless Wrongs in Moral Cognition: Automatic Dyadic Completion from Sin to Suffering." *Journal of Experimental Psychology: General* 143 (4): 1600–1615.
- Gray, Kurt, Adam Waytz, and Liane Young. 2012. "The Moral Dyad: A Fundamental Template Unifying Moral Judgment." *Psychological Inquiry* 23: 206–15.
- Greene, Joshua. 2007. "The Secret Joke of Kant's Soul." In *Moral Psychology*, 3:35–80. Cambridge, MA: MIT Press.
- Greene, Joshua D., Sylvia A. Morelli, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen. 2008. "Cognitive Load Selectively Interferes with Utilitarian Moral Judgment." *Cognition* 107 (3): 1144–54. <https://doi.org/10.1016/j.cognition.2007.11.004>.
- Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen. 2004. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron* 44 (2): 389–400. <https://doi.org/10.1016/j.neuron.2004.09.027>.

- Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (5537): 2105–8. <https://doi.org/10.1126/science.1062872>.
- Haidt, Jonathan, Jesse Graham, and Peter H. Ditto. 2015. "A Straw Man Can Never Beat a Shapeshifter: Response to Schein and Gray (2015)." *YourMorals Blog* (blog). October 27, 2015. <http://www.yourmorals.org/blog/2015/10/a-straw-man-can-never-beat-a-shapeshifter/>.
- Haidt, Jonathan, and Craig Joseph. 2007. "The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules." In *The Innate Mind*, edited by Peter Carruthers, Stephen Laurence, and Stephen Stich. Vol. 3. Oxford: Oxford University Press.
- Hamlin, J. Kiley. 2012. "A Developmental Perspective on the Moral Dyad." *Psychological Inquiry* 23 (2): 166–71. <https://doi.org/10.1080/1047840X.2012.670101>.
- Hamlin, J. Kiley, Karen Wynn, and Paul Bloom. 2007. "Social Evaluation by Preverbal Infants." *Nature* 450 (7169): 557–59.
- Hare, R.M. 1972. "Nothing Matters." In *Applications of Moral Philosophy*. Macmillan. ———. 1982. *Moral Thinking: Its Levels, Methods, and Point*. New York: Oxford University Press.
- Harman, Gilbert. 1999. "Moral Psychology Meets Social Psychology." *Proceedings of the Aristotelian Society* 99: 315–31.
- Hauser, Mark. 2006. *Moral Minds: How Nature Designed a Universal Sense of Right and Wrong*. New York: HarperCollins.
- Hobbes, Thomas. 1998. *Leviathan*. Oxford: Oxford University Press.
- Hofmann, Wilhelm, Daniel C. Wisneski, Mark J. Brandt, and Linda J. Skitka. 2014. "Morality in Everyday Life." *Science* 345 (6202): 1340. <https://doi.org/10.1126/science.1251560>.
- Homer. 2017. *The Odyssey*. Translated by Emily Wilson. W.W. Norton & Company.
- Hooker, Brad. 2015. "Rule Consequentialism." *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/consequentialism-rule/>.
- Huebner, Bryce, James J. Lee, and Marc D. Hauser. 2010. "The Moral-Conventional Distinction in Mature Moral Competence." *Journal of Cognition and Culture* 10 (1–2): 1–26. <https://doi.org/10.1163/156853710X497149>.
- Huemer, Michael. 2005. *Ethical Intuitionism*. New York: Palgrave MacMillan.
- Hurka, Thomas. 1993. *Perfectionism*. Oxford: Oxford University Press. ———. 2006. "Virtuous Act, Virtuous Disposition." *Analysis* 66: 69–76.
- Hursthouse, Rosalind. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.
- Jamieson, Dale, and Robert Elliot. 2009. "Progressive Consequentialism." *Philosophical Perspectives* 23: 241–51.
- Jayawickreme, Eranda, and William Fleeson. 2017. "Does Whole Trait Theory Work for the Virtues?" In *Moral Psychology, Volume 5: Virtue and Character*, edited by Walter Sinnott-Armstrong and Christian Miller, 75–104. Cambridge, MA: MIT Press.
- Joyce, Richard. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press. ———. 2007. *The Evolution of Morality*. Cambridge, MA: MIT Press.

- Kalderon, Mark. 2005. *Moral Fictionalism*. Oxford: Clarendon Press.
- Kamtekar, Rachana. 2004. "Situationism and Virtue Ethics on the Content of Our Character." *Ethics* 114 (3): 458–91.
- Kim, Jaegwon. 2003. "The American Origins of Philosophical Naturalism." *Journal of Philosophical Research* APA Centennial Volume: 83–98.
- Kitcher, Phillip. 2011. *The Ethical Project*. Cambridge, MA: Harvard University Press.
- . 2012. "Naturalistic Ethics Without Fallacies." In *Evolution and Morality*, edited by James E. Fleming and Sanford Levinson, 3–30. New York: NYU Press.
- Kohlberg, Lawrence. 1969. "Stage and Sequence: The Cognitive-Developmental Approach to Socialization." In *Handbook of Socialization and Research*, by David A. Goslin, 347–480. Chicago: Rand McNally.
- Kuhn, Thomas S. 2012. *The Structure of Scientific Revolutions*. 4th ed. Chicago: University of Chicago Press.
- Kurth, Charlie. 2017. "Anti-Realist Moral Progress and the Problem of Moral Reformers." presented at the American Philosophical Association - Pacific Division Meeting, Seattle, WA.
- Locke, John. 2010. *John Locke: A Letter Concerning Toleration and Other Writings*. Indianapolis: Liberty Fund.
- Lutz, Matthew, and James Lenman. 2018. "Moral Naturalism." *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/naturalism-moral>.
- Machery, Edouard. 2010. "The Bleak Implications of Moral Psychology." *Neuroethics* 3 (3): 223–31.
- Mackie, John L. 1991. *Ethics: Inventing Right and Wrong*. Penguin Books.
- Markovits, Julia. 2010. "Acting for the Right Reasons." *The Philosophical Review* 119 (2): 2014–2242.
- McDowell, John. 1995. "Two Sorts of Naturalism." In *Virtues and Reasons: Philippa Foot and Moral Theory*, edited by Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn. Oxford: Oxford University Press.
- . 1998. *Mind, Values, and Reality*. Cambridge, MA: Harvard University Press.
- Mead, Margaret. 1961. "Some Anthropological Considerations Concerning Natural Law." *Natural Law Forum* 6.
- Merritt, Maria. 2000. "Virtue Ethics and Situationist Personality Psychology." *Ethical Theory and Moral Practice* 3 (4): 365–83.
- Merritt, Maria, John Doris, and Gilbert Harman. 2010. "Character." In *The Moral Psychology Handbook*, edited by John Doris, 355–401. Oxford: Oxford University Press.
- Michsel, Walter. 1999. "Personality Coherence and Dispositions in a Cognitive-Affective Personality System (CAPS) Approach." In *The Coherence of Personality: Social-Cognitive Bases of Consistency, Variability, and Organization*, edited by Daniel Cervone and Yuichi Shoda, 37–60. New York: Guilford Press.
- Mikhail, John. 2007. "Universal Moral Grammar: Theory, Evidence and the Future." *Trends in Cognitive Sciences* 11 (4): 143–52. <https://doi.org/10.1016/j.tics.2006.12.007>.

- . 2009. “Is the Prohibition of Homicide Universal? Evidence from Comparative Criminal Law.” *Brooklyn Law Review* 75: 497–515.
- Mill, John Stuart. 1974. *On Liberty*. Harmondsworth: Penguin.
- . 1979. *Utilitarianism*. Edited by George Sher. Indianapolis: Hackett.
- Miller, Alexander. 2013. *Contemporary Metaethics: An Introduction*. 2nd ed. Malden, MA: Polity Press.
- Miller, Christian. 2013. *Moral Character: An Empirical Theory*. Oxford: Oxford University Press.
- Miller, Christian B. 2016. “Empirical Approaches to Moral Character.” *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/moral-character-empirical/#CAPSMODE>.
- Miller, Geoffrey F. 2007. “Sexual Selection for Moral Virtues.” *The Quarterly Review of Biology* 82 (2): 97–125.
- Mischel, Walter, and Yuichi Shoda. 1995. “A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure.” *Psychological Review* 102 (2): 246–48.
- Moore, G.E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.
- Morton, Adam. 2013. *Bounded Thinking: Intellectual Virtues for Limited Agents*. Oxford: Oxford University Press.
- Nagel, Thomas. 1971. “The Absurd.” *The Journal of Philosophy* 68 (20): 716–27.
- . 2012. *Mind and Cosmos*. Oxford: Oxford University Press.
- Noddings, Nel. 2003. *Caring: A Feminine Approach to Ethics and Moral Education*. 2nd ed. University of California Press.
- Norcross, Alastair. 1997. “Good and Bad Actions.” *Philosophical Review* 106 (1): 1–34.
- Nussbaum, Martha C. 2000. *Women and Human Development: The Capabilities Approach*. New York: Cambridge University Press.
- Ohtsubo, Yohsuke, and Esuka Watanabe. 2008. “Do Sincere Apologies Need to Be Costly? Test of a Costly Signaling Model of Apology.” *Evolution and Human Behavior* 30 (2): 114–23. <https://doi.org/10.1016/j.evolhumbehav.2008.09.004>.
- Papineau, David. 2015. “Naturalism.” *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/naturalism/>.
- Parfit, Derek. 2013. *On What Matters*. Oxford: Oxford University Press.
- Pereboom, Derk. 2006. *Living Without Free Will*. Cambridge Studies in Philosophy. Cambridge, UK: Cambridge University Press.
- Peterson, Martin. 2013. *The Dimensions of Consequentialism*. Cambridge: Cambridge University Press.
- Pettit, Philip. 2015. *The Robust Demands of the Good: Ethics with Attachment, Virtue, and Respect*. New York: Oxford University Press.
- . 2018. “Three Mistakes about Doing Good (and Bad).” *Journal of Applied Philosophy* 31 (1): 1–25.
- Phillips, D.Z. 2005. *The Problem of Evil and the Problem of God*. Fortress Press.
- Piaget, Jean. 1965. *The Moral Judgment of the Child*. Translated by Marjorie Gabain. New York: Free Press.
- Pizarro, David, A., and David Tannebaum. 2012. “Bringing Character Back: How the Motivation to Evaluate Character Influences Judgments of Moral Blame.” In

- The Social Psychology of Morality: Exploring the Causes of Good and Evil*, edited by Mario Mikulincer and Phillip R. Shaver, 91–108. Washington, DC: American Psychological Association.
- Pizarro, David, A., David Tannebaum, and Eric Uhlmann. 2012. “Mindless, Harmless, and Blameworthy.” *Psychological Inquiry*, no. 23: 1–4.
- Popper, Karl. 2002. *The Logic of Scientific Discovery*. Routledge.
- Railton, Peter. 1984. “Alienation, Consequentialism, and the Demands of Morality.” *Philosophy and Public Affairs* 13: 134–71.
- Rawls, John. 1980. “Kantian Constructivism in Moral Theory: The Dewey Lectures 1990.” *Journal of Philosophy* 77 (9): 515–72.
- Reeder, Glenn D., and Marilyn B. Brewer. 1979. “A Schematic Model of Dispositional Attribution in Interpersonal Perception.” *Psychological Review* 86 (1): 61–79.
- Ridge, Michael. 2014. “Moral Non-Naturalism.” *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/moral-non-naturalism/>.
- Riggs, Wayne D. 2016. “Open-Mindedness, Insight, and Understanding.” In *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, edited by Jason Baehr, 18–37. Routledge.
- Robertson, Seth. 2018. “Power, Situation, and Character: A Confucian-Inspired Response to Indirect Situationist Critiques.” *Ethical Theory and Moral Practice* 21 (2): 341–58.
- Ross, David. 1930. *The Right and The Good*. Oxford: Clarendon Press.
- Rothfuss, Patrick. 2008. *The Name of the Wind*. DAW Books.
- Rottman, Joshua, Liane Young, and Deborah Kelemen. 2017. “The Impact of Testimony on Children’s Moralization of Novel Actions.” *Emotion* 17 (5): 811–27. <https://doi.org/10.1037/emo0000276>.
- “Roundabout Benefits.” n.d. The Washington State Department of Transportation. <https://www.wsdot.wa.gov/Safety/roundabouts/benefits.htm>.
- Ruse, Michael, and E.O. Wilson. 1995. “Moral Philosophy as Applied Science.” In *Conceptual Issues in Evolutionary Biology*, edited by Elliot Sober, 2nd ed., 421–38. Cambridge, MA: MIT Press.
- Russell, Daniel. 2009. *Practical Intelligence and the Virtues*. Oxford: Oxford University Press.
- Sarkissian, Hagop. 2010. “Confucius and the Effortless Life of Virtue.” *History of Philosophy Quarterly* 27 (1): 1–16.
- Scanlon, Thomas M. 1998. *What We Owe To Each Other*. Cambridge, MA: Harvard University Press.
- Schein, Chelsea, Amelia Goranson, and Kurt Gray. 2015. “The Uncensored Truth about Morality.” *The Psychologist* 28 (12): 982–85.
- Schein, Chelsea, and Kurt Gray. 2015. “The Unifying Moral Dyad: Liberals and Conservatives Share the Same Harm-Based Moral Tempalte.” *Personality and Social Psychology Bulletin* 41 (8): 1147–63.
- . 2017. “The Theory of Dyadic Morality: Reinventing Moral Judgment by Redefining Harm.” *Personality and Social Psychology Review*, 1–39.
- Schüll, Natasha Dow. 2014. *Addiction by Design: Machine Gambling in Las Vegas*. Princeton: Princeton University Press.

- Sen, Amartya. 1979. "Utilitarianism and Welfarism." *Journal of Philosophy* 76: 463–89.
- Shafer-Landau, Russ. 2005. *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Sheskin, Mark, and Laurie Santos. 2012. "The Evolution of Morality: Which Aspects of Human Moral Concerns Are Shared With Nonhuman Primates?" In *The Oxford Handbook of Comparative Evolutionary Psychology*, edited by Todd K. Shackelford and Jennifer Vonk. Oxford: Oxford University Press.
<http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199738182.001.0001/oxfordhb-9780199738182-e-23>.
- Shweder, Richard. 2012. "Relativism and Universalism." In *A Companion to Moral Anthropology*, 85–102. Hoboken, NJ: Wiley.
- Sidgwick, Henry. 1981. *The Methods of Ethics*. Indianapolis, IN: Hackett Publishing Company.
- Singer, Peter. 1975. *Animal Liberation*. HarperCollins.
- . 1999. "The Singer Solution to World Poverty." *The New York Times Magazine*, 60–36.
- Sinnott-Armstrong, Walter. 2015. "Consequentialism." *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/consequentialism/>.
- Slingerland, Edward. 2011. "The Situationist Critique and Early Confucian Virtue Ethics." *Ethics* 121 (2): 390–419.
- Slote, Michael. 1984. "Satisficing Consequentialism." *Proceedings of the Aristotelian Society, Supplementary Volumes* 58: 139–63.
- . 2001. *Morals from Motives*. Oxford: Oxford University Press.
- Smart, J.J.C. 1956. "Extreme and Restricted Utilitarianism." *The Philosophical Quarterly* 6 (25): 344–354.
- Smart, J.J.C., and Bernard Williams. 1973. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.
- Smilanksy, Saul. 2002. *Free Will and Illusion*. Oxford: Oxford University Press.
- Snow, Nancy. 2009. *Virtue as Social Intelligence: An Empirically Grounded Theory*. Routledge.
- Snyder, Melvin L., Robert Kleck, Angelo Strenta, and Steven J. Mentzer. 1979. "Avoidance of the Handicapped: An Attributional Ambiguity Analysis." *Journal of Personality and Social Psychology* 37: 2297–2306.
<https://doi.org/10.1037//0022-3514.37.12.2297>.
- Sreenivasan, Gopal. 2013. "The Situationist Critique of Virtue Ethics." In *The Cambridge Companion to Virtue Ethics*, edited by Daniel Russell, 290–314. Cambridge: Cambridge University Press.
- . 2017. "Character Education and the Rearguard of Situationism." In *Moral Psychology, Vol 5, Virtue and Character*, edited by Walter Sinnott-Armstrong and Christian Miller, 131–63. Cambridge, MA: MIT Press.
- Stohr, Karen. 2018. "Virtuous Motivation." In *The Oxford Handbook of Virtue*, 453–69. Oxford: Oxford University Press.
- Stratton-Lake, Philip. 2014. "Intuitionism in Ethics." *Stanford Encyclopedia of Philosophy*.

- Strawson, Galen. 1994. "The Impossibility of Moral Responsibility." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 75 (1/2): 5–24.
- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127: 109–66.
- . 2010. "What Is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5 (5): 363–84.
- Stump, Eleonore. 2010. *Wandering in Darkness: Narrative and the Problem of Suffering*. Oxford: Oxford University Press.
- Sturgeon, Nicholas. 1986. "Harman on Moral Explanations of Natural Facts." *Southern Journal of Philosophy* 24: Supplement.
- Surin, Kenneth. 1986. *Theology and the Problem of Evil*. Oxford: Blackwell.
- Sverklik, Steven. 2011. *Motive and Rightness*. Oxford: Oxford University Press.
- The Economist*. 2013. "The Theology of Self-Destruction," March 22, 2013. <https://www.economist.com/erasmus/2013/03/22/the-theology-of-self-destruction>.
- "The Frankfurt Kitchen." n.d. *The Museum of Modern Art* (blog). https://www.moma.org/interactives/exhibitions/2010/counter_space/the_frankfurt_kitchen.
- Thompson, Andrew. 2015. "Engineers of Addiction." *The Verge*, 2015. <https://www.theverge.com/2015/5/6/8544303/casino-slot-machine-gambling-addiction-psychology-mobile-games>.
- Thompson, Michael. 2008. *Life and Action: Elementary Structures of Thought and Practice*. Cambridge, MA: Harvard University Press.
- Tilley, Terrence W. 1991. *The Evils of Theodicy*. Washington, DC: Georgetown University Press.
- Trivers, Robert L. 1971. "The Evolution of Reciprocal Altruism." *The Quarterly Review of Biology* 46 (1): 33–57.
- Turiel, Elliot, Melanie Killen, and Charles C. Helwig. 1987. "Morality: Its Structure, Functions, and Vagaries." In *The Emergence of Morality in Young Children.*, 155–243. Chicago, IL, US: University of Chicago Press.
- Uhlmann, Eric, David Pizarro A., and Daniel Diermeier. 2015. "A Person-Centered Approach to Moral Judgment." *Perspectives on Psychological Science* 10 (1): 72–81.
- Unger, Peter. 1996. *Living High and Letting Die: Our Illusion of Innocence*. Oxford: Oxford University Press.
- Van Buren, Martin. 1867. *Inquiry Into the Origin and Course of Political Parties in the United States*. New York: Hurd and Houghton.
- Van de Vondervoort, Julia, and J Kiley Hamlin. 2016. *Evidence for Intuitive Morality: Preverbal Infants Make Sociomoral Evaluations*. <https://doi.org/10.1111/cdep.12175>.
- Vargas, Manuel R. 2013. *Building Better Beings: A Theory of Moral Responsibility*. Oxford: Oxford University Press.
- Vavova, Katia. 2015. "Evolutionary Debunking of Moral Realism." *Philosophy Compass* 10 (2): 104–16.

- Verplaetse, Jan, Sven Vanneste, and Johan Braeckman. 2007. "You Can Judge a Book by Its Cover: The Sequel. A Kernel of Truth in Predictive Cheating Detection." *Evolution and Human Behavior* 28: 629–47.
<https://doi.org/10.1016/j.evolhumbehav.2007.04.006>.
- Wiggins, David. 1991. "Moral Cognitivism, Moral Realitivism, and Motivating Moral Beliefs." *Proceedings of the Aristotelian Society* 91: 61–85.
- . 1998. *Needs, Values, Truth*. Third. Oxford: Clarendon Press.
- Wolf, Susan. 1982. "Moral Saints." *Journal of Philosophy* 79 (8): 419–39.
- Zagzebski, Linda. 2010. "Exemplarist Virtue Theory." *Metaphilosophy* 41 (1–2): 41–57.
- . 2017. *Exemplarist Moral Theory*. New York: Oxford University Press.

[Appendix A: Truth, Relativism, and Pluralism]

Here, I would like to briefly address a few issues concerning perspectival naturalism, the metaethical view I developed in this dissertation. While I discussed perspectival naturalism at length in Chapter 3, these three issues are best addressed after we've seen the view developed in greater detail and applied to the argument for the Consequences and Character Conditions. My discussion here will be brief: I can only provide sketches, to be filled out in more detail in the future, of responses to these issues. The first issue is expository. I claimed that perspectival naturalism is more interested in the justification of moral systems than in the truth of moral claims – but I only mentioned in passing what moral truth amounts to for perspectival naturalism. And so, one might ask for more details here: according to perspectival naturalism, are there moral truths, and if so, what makes them true? The next two issues are important objections to perspectival naturalism. The first objection is that perspectival naturalism inherits from its constructivist roots a deeply problematic circularity: it claims that what makes things valuable is, ultimately, that we value them. So, can perspectival naturalism provide a non-circular account of the foundation of moral justification? Third, perspectival naturalism is explicitly committed to a kind of pluralism about moral systems. It claims that we should always treat our moral systems as open to revision (this was the burden shifting argument in Chapter 3), and that different moral systems could be better or worse designed for different groups of people at different times, based on shared circumstances, histories, and cultural meanings (think here of the example of *xenia*). But, can perspectival naturalism defend this pluralism without reverting into radical relativism?

Moral Truth

Let's address these issues one by one. The first issue is the question of what moral truth amounts to for perspectival naturalism. The short answer to this question is that, for perspectival naturalism, moral truths are truths about the best-designed moral systems for us. Thus, for a moral claim such as "murder is wrong" to be true, it must be entailed from features of the best designed moral systems for us. But, this raises an immediate objection: we do not yet know what the best designed moral systems are for us, and so it seems that we must say that we do not yet know the truth of any moral claim. Further, since I argued that our moral systems are essentially open to revision, I seem further committed to the even moral radical claim that we *could not* know the truth of any moral claim. But, this is too hasty: things are not so dire as they appear. Admittedly, on this view, we are not ever justified in being absolutely certain in the truth of a full moral system or its particular claims, but that is exactly how it should be. This is not a particularly unique feature of perspectival naturalism: when viewed from a philosophical or skeptical perspective, we ought not be absolutely certain of any claim, whether moral or not. Of course, the best-designed moral systems will likely suggest that we refrain from spending all of our time in philosophical or skeptical perspectives, and in lived perspectives, there are compelling reasons to be certain of some things. The general point, however, is that in real life, we are always making our best guesses: if I say I am certain that Paris is the capital of France, or that the softball baserunner was clearly safe, there is always some chance that I could be incorrect (even if this chance comes from a radical skeptical scenario like an evil demon or being a brain in a vat). This fact doesn't bother us terribly much in other scenarios and circumstances, so it

shouldn't bother us terribly much when it comes to perspectival naturalism. There is a more pointed version of this concern however: why should we be confident that the moral systems we employ in general are remotely close to the best-designed moral systems for us? If our moral commitments are *close enough* to the best-designed moral systems for us, then we need not worry too much, but if the best-designed moral systems for us are radically different from those we currently employ, then we *should* doubt our current moral systems and the claims entailed therein. Fortunately, there are at least two reasons we should not radically doubt our current moral systems. First, if my claim is correct that morality has naturally been undergoing just the process of improvement that perspectival naturalism suggests, then we have already done quite a bit of work towards closing in on well-designed theories. Second, well-designed moral systems need to meet people where they are: they need to work based on the circumstances, traditions, shared histories and patterns of meaning, understanding, and interpretation, and everything else that we bring to the table with us. At a given time in given circumstances, different moral systems are better or worse designed for us, and the best-designed systems that we have available, even if they are not well-enough designed to ground moral truth still ground normativity: until better systems are designed, we should follow the best of what we have.

Foundations

The second concern I will address here focuses on whether or not perspectival naturalism's account of moral-theoretical justification is circular. To see how such an objection might go, consider John Stuart Mill's famous "proof" in Chapter 4 of

Utilitarianism that only happiness is ultimately desirable (or, as we might say, valuable).

He claims:

The only proof capable of being given that an object is visible is that people actually see it. The only proof that a sound is audible is that people hear it; and so of the other sources of our experience. In a like manner, I apprehend, the sole evidence it is possible to produce that anything is desirable is that people do actually desire it. (4.3)

And shortly after:

[...] if human nature is so constituted as to desire nothing which is not either a part of happiness or a means of happiness, we can have no other proof, and we require no other, that these are the only things desirable. (4.9)

G.E. Moore (who was not always the most charitable reader of late 19th century British ethicists) wrote in response that Mill made a mistake:

...so obvious, that it is quite wonderful how Mill failed to see it. The fact is that "desirable" does not mean "able to be desired" as "visible" means "able to be seen." The desirable means simply what ought to be desired or what deserves to be desired. (p 53)

Unsurprisingly, Sidgwick had also (previously) noticed this apparently odd move:

...in giving as a statement of this principle that "the general happiness is *desirable*," [Mill] must be understood to mean (and his whole treatise shows that he does mean) that is it what each individual *ought* to desire...But this proposition is not established by Mill's reasoning, even if we grant that what is actually desired may be legitimately inferred to be in this sense desirable. (388, III, xiii, 4)

Sidgwick highlights in a footnote to this passage that he was here thinking specifically of a "confusion in Mill's mind between two possible meanings of the term 'desirable,' (1) what can be desired and (2) what ought to be desired" (p. 388, ff2).

Now, one might be tempted to say that Mill was just, somehow, unaware of the type of mistake that he was making – that he was not clear in general on the distinction between descriptive claims and normative claims. But, this cannot be correct – he discusses the distinction with clarity in his popular logic textbook:

Every art has one first principle, or general major premise, not borrowed from science; that which enunciates the object aimed at, and affirms it to be a desirable object. The builder's art assumes that it is desirable to have buildings; architecture, as one of the fine arts, that it is desirable to have them beautiful or imposing. The hygienic and medical arts assume, the one that the preservation of health, the other that the cure of disease, are fitting and desirable ends. These are not propositions of science. Propositions of science assert a matter of fact: an existence, a co-existence, a succession, or a resemblance. The propositions now spoken of do not assert that any thing is, but enjoin or recommend that something should be. They are a class by themselves. A proposition of which the predicate is expressed by the words *ought* or *should be*, is generically different from one which is expressed by *is*, or *will be*. (VI, XII, 6; italics in original)

So, Mill was clearly aware in general that there is a *prima facie* distinction between the claim that something is actually desired and something ought to be desired – yet he certainly appears to collapse the distinction in the *Utilitarianism* passage. There have been several attempts to save Mill from having made such an apparently bad mistake, and they tend to argue that given Mill's other theoretical commitments, such as a commitment to a fairly radical version of empiricism, he could not make any claim that things were desirable outside of the observable fact that people desired them. And, we will take a somewhat similar line here for perspectival naturalism.

According to perspectival naturalism, justification for moral claims (and perspectives, and systems) cannot derive from completely mind-independent moral facts. There are no genuinely categorical imperatives: the potential variety of moral minds is far too staggeringly immense. Thus, if moral justification is to derive from anywhere, it must derive ultimately from values (broadly construed) that we do or could hold. But still, we do not want to fall into the trap that Mill might (or might not) have: we don't want to claim that what is (ultimately, or in the best moral system) morally valuable is simply what we (antecedently) morally value. Thus, we need to be careful

here to distinguish between two distinct questions: what justifies treating certain considerations as *candidates* for moral considerations, and what justifies treating certain considerations as *ultimate* moral considerations? The first question gives us a base group of potential moral considerations for a moral system. The second question gives us the actual moral system we should adopt. And thus, we can see that for the first question, Mill's line of reasoning is perfectly reasonable. For something to be a genuine candidate moral consideration for us (e.g. the welfare of other people or myself, happiness, flourishing, human dignity, honesty, religious piety), it must be an expression of or ultimately derive from our values and preferences. Now, Mill went wrong in claiming that all candidates for moral considerations shared the feature of being ultimately aimed at happiness or utility: there are many types of moral considerations along many different trajectories and they do not all ultimately aim at any single thing. If they did, moral philosophy would be a great deal easier. Instead, we are left with the project of *managing* these candidate considerations *using* them.

Pluralism and Relativism

In describing perspectival naturalism, I noted several defining features of the view. For example, the view was committed to moral anti-realism, and to the Design paradigm. One commitment I did not discuss in any great detail is that perspectival naturalism is essentially pluralistic. This follows from the claim that, for perspectival naturalism, our moral systems should always remain open to revision. Thus, our moral systems (following that argument, laid out in Chapter 3), should be pluralistic in the sense that we should always accept that (a) in different circumstances, slightly different moral systems might be better designed than those that are well-designed for our current

circumstances, (b) even in our current circumstances, there might be better-designed moral systems because no moral system is perfect (in the sense of the term used in the Chapter 3 argument). Further, there is a plausible design argument in favor of moral systems that treat pluralism (in the sense of a pluralistic tolerance for minimally acceptable) varying moral, political, and religious worldviews. This just is the argument that has been given in various forms since Locke's *A Letter Concerning Toleration* and Mill's discussion of toleration in *On Liberty*: that given human beings' propensity for severe and horrific violence towards outgroup members, toleration is a vital tool for maintaining people's welfare.

But, all this raises a challenge for perspectival naturalism. Can perspectival naturalism be pluralistic without reverting to a radical form of relativism? To make the question more concrete, how could perspectival naturalism justify the claim that certain acts like genocide or child abuse are morally wrong across any culture? We should start by answering this question at a very general level, and point out that for any act, if there is a compelling case to be made that we should treat it as wrong in all cultures, then a well-designed moral system would be responsive to that case. But, this point risks trivializing the view: it says that if radical relativism (or anything else) is incorrect, then perspectival naturalism would treat it as such. Fortunately, we can give a more robust answer here, but with a caveat. We must remember the general Humean constructivist point that morality in general *is* relative to creatures like us – to our circumstances and needs, and the things that we do and can value and disvalue. So, we need to temper our expectations here: as I emphasized throughout Chapter 3, there is no mind-independent realm of moral facts that we can appeal to if we wish to avoid radical relativism. We

must make do with what we have. Thus, my more robust answer to this challenge is based largely on the discussion in Section 4.2.1 of the “dyadic loop”: the psychological tendency for our evaluations of harm and our evaluations of moral wrong to be mutually reinforcing and (thus) for considerations related to harm and welfare to be especially invasive in moral deliberation. Call “compassion” the moral consideration for being concerned about noticing, mitigating, and repairing the types of harm most salient in the dyadic loop: intentional agents harming vulnerable patients. Now, we can argue that well-designed moral systems for us should treat compassion as central or foundational: not because it is philosophically or explanatorily *foundational* but because it is philosophically or explanatorily *invasive*: harm-based considerations, if empirical claims about the dyadic loop are correct, will always pop up in moral deliberation sooner or later. If we do treat harm and compassion as central or foundational for well-designed moral systems, we should note two immediate consequences. First, this gives us a way forward in certain cases of moral disagreement in which the other relevant moral considerations can be weighed without settling upon an answer. Second, and more importantly for us here, this point helps us respond to this worry about radical relativism. Notice that, in most of the egregious cases that give us pause concerning radical relativism, just the type of harm described in the dyadic loop is especially salient. If compassion is central or foundational in well-designed moral systems, then we have, at least in part, a robust answer to this relativist challenge.