ON REASONS TO LIVE JUSTIFIABLY:
IN SUPPORT OF A HUMEAN CONTRACTUALIST ACCOUNT OF MORAL REASONS

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Philosophy

By

Travis N. Rieder, M.A.

Washington, DC
May 5, 2014
ON REASONS TO LIVE JUSTIFIABLY:
IN SUPPORT OF A HUMEAN CONTRACTUALIST ACCOUNT OF MORAL REASONS

Travis N. Rieder, M.A.

Thesis Advisor: Henry S. Richardson, J.D. Ph.D.

ABSTRACT

The goal of this dissertation is to explore a new answer to the very old question, “Why be moral?” Or, as the question is often phrased today, “What reason does one have to be moral?” I begin my investigation into this question (what I call the ‘WM’ question) with the keen analysis of it given by T. M. Scanlon in articulating his contractualist answer in What We Owe to Each Other (1998). Although I take Scanlon’s view – what he calls an account of the ‘motivational basis of morality’ – to be quite promising, I lodge two objections against it. First, Scanlon does not avail himself of the most complete answer that he could to the WM question. And second, his account of which reason one has to be moral is imbedded in a grand theory that provides a problematic explanation of the existence of reasons. I then argue that strengthening Scanlon’s account in response to the first observation makes attractive an alternative theory of reasons, the adoption of which responds to the second objection. In short, I argue that Scanlon can and should accept an account on which one has reason to act rightly because one has reason to live with others on mutually justifiable terms – a reason that one has because she has the further reason to have intimate relationships with others. This latter reason, then, is explainable in a plausible, naturalistic way by the Humean Theory of Reasons (HTR), on which an agent’s particular desires explain her reasons for action.
The result of modifying Scanlon’s account in the direction suggested is a view that I call a Humean Contractualist Account of Moral Reasons, or HCA. After making the initial case for such a view in the first two chapters, I spend chapters 3 and 4 completing a sketch of HCA. I then spend the final three chapters defending it from important criticisms. In the end, I conclude that HCA is promising and deserving of further consideration.
What I now think of as the first draft of this dissertation was written nearly eight years ago, as a term paper in my first semester as a graduate student. Between then and now, I have had an unbelievable amount of support from my mentors, colleagues, friends and family. As such, I don’t even know how to begin thanking them all. So I will simply do my best, knowing that I likely have left many people out. But they all have my gratitude.

To Justin Weinberg, for encouraging the early, fledgling drafts of this project, and for introducing me to my eventual mentor, Henry Richardson. And to Kevin Elliott for taking the ideas articulated here seriously, and for believing that some of the things I argue for may even be true.

To my friends and colleagues – at the University of South Carolina, at Georgetown, and at the various conferences and workshops I attended over the years – who have read and listened to many of the arguments presented here, and who have provided helpful thoughts, insightful criticism, and wonderful discussions.

To my dissertation Readers, Mark Murphy and Steve Kuhn, who caught many errors, and provided me with more feedback than I could have been justified in requesting; and to my Advisor, Henry Richardson, from whom I learned how to do professional philosophy. Whatever care, precision and rigor appear in these pages is due largely to the influence of these three individuals.

And finally: to my parents, Joyce and Brent, for raising me to be curious; to Yumosh, for keeping me company while I wrote; to my amazing partner, Sadiye, for her love and support; and to our beautiful daughter, Sinem, whose arrival motivated the completion of this dissertation, and whose smiles kept me sane in the meantime.

To all of these people, and for all of these reasons, I am indebted.

With great appreciation,
Travis N. Rieder
# Table of Contents

Introduction............................................................................................................................. 1

Chapter 1.................................................................................................................................. 16

Chapter 2................................................................................................................................... 49

Chapter 3................................................................................................................................... 84

Chapter 4................................................................................................................................... 126

Chapter 5................................................................................................................................... 160

Chapter 6................................................................................................................................... 209

Conclusion .................................................................................................................................. 243

Bibliography ............................................................................................................................... 255
INTRODUCTION

The central question of this dissertation is the very old one, “Why be moral?” Or, as I tend to think of it now: “What reason does one have to be moral?” As an undergraduate at a small, liberal-arts college, where heavy, late-night philosophical discussions were common, I remember being especially bothered by Dostoyevsky’s version of this challenge, which focuses on what the answer could be without an all-powerful, rewarding or punishing god. If such a god doesn’t exist, what reason could one have to be moral? Isn’t it the case that, if God doesn’t exist, ‘everything is permitted’? I quickly came to believe that the existence or non-existence of a god would not, in fact, help or hinder my attempt to answer this question (at least, not in the way my 20-year-old brain had expected), since promise of eternal reward or punishment did not offer a satisfactory answer anyway. And so, as I continued my philosophical education, I kept looking for answers – a search which ultimately led to this dissertation.

Over the course of the next seven chapters, I will slowly build and defend a new answer to the ‘why be moral’ question, or as I will later call it, the ‘WM’ question. However, as I will explain further below, to say that I ‘argue for’ an answer to the WM question would be a bit misleading; although I certainly provide arguments for aspects of the view I lay out, I do not take myself to be providing anyone with sufficient reason to take my answer to be the answer. Instead, I am here engaging simultaneously in two, more modest projects: first, I am answering the question of what is plausibly true concerning one’s reason to be moral, whatever else is true; and second, I am offering my answer to that first question as a view of moral reasons that deserves further attention as a candidate for the correct view. In what follows, I will say a bit
more about this methodology, after which I will preview the key arguments and positions to come.

0.1 The Methodology of the Dissertation

Long after those late nights in college, I came to identify my interest in a non-religious answer to the WM question as being part of an overall philosophical modesty to which I was committed. This modesty sometimes comes out as an expression of ‘naturalism’, ‘minimalism’, or even ‘eliminativism’ in various discussions, but I do not here want to defend any specific one of these positions. Instead, I will just stick with the language of modesty to signify an interest in philosophical arguments and positions that require as little as possible in the way of substantive philosophical assumptions. On this understanding, the goal of this dissertation is to see what can be done in terms of answering the WM question modestly.

My hypothesis is that the answer is: more than one might expect. I tend to think that our (philosophers’) knee-jerk reactions to views and positions that seem to get something wrong harms our ability to clearly evaluate philosophical positions. In the case of explaining moral reasons, for instance, some views have the (admittedly) strange implication that hypothetical, idiosyncratic characters have no reason to be moral. An example, explored in detail in Chapter

---

1 As an example: although many will identify my project here as a naturalistic one, I am not arguing, as a naturalist might, that one reason to accept the view is its naturalism. Rather, to the extent that the view is appropriately described as naturalistic, the reason for this is that I am attempting to utilize positions and arguments that most anyone could accept. And so, while the non-naturalist, for instance, does not believe that the natural exhausts what there is, the non-naturalists does believe in the natural. Thus, developing a theory that is broadly naturalistic is part of my methodology, not because I want to argue for naturalism, but because I want to develop a position that anyone could accept, whether or not they might in fact accept other positions as well.
6, is that of Alan Gibbard’s ‘ideally coherent Caligula’. According to some accounts of the reason to be moral (including my own), if there were such a person as Caligula, who had a perfectly coherent psychology (although not logically impossible, likely practically impossible), then this individual would have no reason not to, say, torture for fun. And a common reaction to such an implication is as follows: well, it is obvious that Caligula, coherent or not, has a reason not to torture for fun, and so the view under consideration must be false. It is, as we like to say, a *reductio* of the view.

I tend to think that many fewer things are ‘*reductios*’ than are claimed. The better strategy for evaluating the success of any position, whether it implies that the Ideally Coherent Caligula has a reason to be moral or not, is to explore the account in some detail and to spend a good deal of time asking just how successful the view is, overall, according to many theoretical dimensions. And when we do this, I suspect that many fewer views are ‘obviously false’ than are claimed to be so.

Back to the hypothesis, then: answering the WM question modestly can accomplish more than one might expect. I am not hypothesizing, notice, that a modest view can do *everything* – which is to say, that it might be explanatorily perfect. Rather, I am predicting that, if we give a modest answer to the WM question a fair shake, we might be surprised at how successful it is.

For my purposes, there are two interesting ways in which the relative success of my answer to the WM question would matter. Most ambitiously, it might be so (surprisingly) successful at explaining one’s reason to be moral, that it provides good reason not to advance a more theoretically ambitious view. That is: once we give the view a fair shake, as it were, we decide that it is so successful by any reasonable standard that, given its additional virtue of
modesty, we decide to treat it as a serious candidate for being the correct view. Indeed, this is my own suspicion about the view that I ultimately defend.

However, as will become obvious, I simply cannot, in this one document, provide the number of arguments that would be required to be considered a serious defense of some particular view of the reason to be moral. Why? Because an explanation of ‘the reason to be moral’ turns out to be theoretically massive, constituted by an account of permissibility, reasons, the weight of reasons, and moral motivation – each of which could be the subject of a book (or many!) on its own. And so, rather than take as my primary aim to argue for my own view of the reason to be moral, I will here claim to put forward a view motivated by its philosophical modesty, but which is also surprisingly successful. That is, I will return, at the end, to the question of whether I have rendered the view articulated ‘defensible’, and suggest that I have; but if it’s true, this will not be a result of ground-up argument for the position.

In contrast to the goal of ‘arguing for’ a view of moral reasons, the primary methodology of the dissertation will aim at the second sense in which my answer to the WM question might succeed: it might, as I put it later, provide a comforting ‘back-up’ theory of moral reasons. A back-up theory is one that might not have some of the virtues of one’s favored primary candidate, but which is there to serve the function in case the primary candidate fails. And why is having a back-up important? Well, because more ambitious theories might be false, and so if we put all of our effort into defending those views, and they fail, then we are left without an explanation of moral reasons. The goal of a modest, back-up view, then, is to provide an account

---

2 I will spend more time discussing this idea explicitly at the end of Chapter 2.
of moral reasons that requires very little buy-in, but which claims to do real philosophical work; any objections to such a view are likely based on its being unsatisfying or incomplete if considered on its own, rather than on its seeming straightforwardly false. A successful articulation of such a view, should it be able to explain at least much of what we want from an account of moral reasons, would then provide a kind of helpful confidence in our claims about moral reasons. Even if one’s favored view turns out to fail, she need not then fear that nihilism about moral reasons is true, or that we can’t accurately make many claims about what we and others have reason to do.

Given that this modest goal will be primary for me, the dissertation will take on a different structure than it otherwise might. I will not argue for every claim or position directly; that is, I will not hope to argue for the truth of each position from unassailable foundations. Instead, I will begin with a view (contractualism) that I take to be plausible as a substantive moral theory, but which has the added virtue of being understandable in a very modest way (more on this below). I will then provide independent reasons to ground contractualism, as I understand it, with an ontologically modest theory of reasons (Humeanism) that is capable of vindicating many of our intuitions, even if it strikes some as incomplete. This marriage of contractualism and the Humean Theory of Reasons forms the groundwork of my attempt to generate a modest theory of moral reasons. Let me now turn to a brief preview of the arguments to be made along the way.

0.2 The Key Arguments and Positions

I call the view offered in this dissertation a Humean Contractualist Account of Moral Reasons, or HCA. While the arguments that support considering such a view require the
expenditure of many, many words, the view itself is quite simple. According to HCA, moral action is justifiable action; one has a reason to act so long as the action promotes some desire that she has; and so one has a reason to act rightly so long as one has a desire that is promoted by acting in justifiable ways. And while this schematic form of the view may not sound all that attractive, I am moved to accept it initially based on a further, substantive view, what I call The Motivational Basis of Morality. This is the view that the desires to live with others on terms that they could accept and to have intimate relationships form the grounds for a complex explanation of many people’s reason to act justifiably. Of course, there are additional moves and arguments along the way, so I will here preview some of the primary ones.

After claiming, in the following section (0.3), that contractualism offers a uniquely attractive starting point for a modest theory of moral reasons, I begin in Chapter 1 by exploring T.M. Scanlon’s important, influential answer to the WM question. There I argue that Scanlon’s own arguments and positions make possible a stronger account of what he calls ‘the motivational basis of morality’ than he, himself, accepts. While Scanlon holds that one has reason to act rightly because she has a reason to live with others on justifiable terms, I argue that this answer is unsatisfying according to Scanlon’s own criteria of success. However, by utilizing others of his arguments, I also am able to show that he could accept (and I think should accept) the ‘complex’ view that one has a reason to act rightly because she has reason to live with others on justifiable terms, and that this reason is further explainable by reference to the reason to stand in intimate relationships with (particular) others. This view, I contend, is entailed by Scanlon’s own positions, and fairs better according to the (compelling) success criteria that he has laid out.
The ‘complex account’ of the reason to be moral, then, sets up my ability to build the rest of HCA. In Chapter 2, I argue that it allows me to reject Scanlon’s problematic ‘primitivism’ about reasons in favor of a ‘Humean’ account of reasons. Since I want to restrict myself, at this point, to a relatively modest argument, I end the chapter by suggesting that one might consider this marriage a ‘backup view’ of moral reasons – that perhaps this is what we could count on being true, even if other, more ambitious, views fail. According to this modest argument, then, what would plausibly be true, even if one’s preferred, more ambitious account was false, is that anyone with a desire for intimacy, or to live justifiably, would have a reason to act rightly.

Of course, at this point, one would be justified in asking how much reason one has to act rightly according to HCA. Answering this question requires providing a theory of the ‘weight’ of reasons, which I attempt to do in Chapter 3. In short, I there argue that, despite a recent, powerful argument by Mark Schroeder to the contrary, a Humean ought to be a ‘proportionalist’ – which is to say that a Humean ought to hold that the weight of a reason is proportional to the strength of the grounding desire, as well as the degree to which acting promotes that desire.

At this point in the argument, the basic structure of HCA is in place: it is constituted by contractualism; HTR; proportionalism; and a substantive, complex account of the motivational basis of morality. In Chapter 4, I step back to pull the view together, and to see what components would ideally need filling in before HCA constituted anything like a complete picture of moral reasons. Although I admit that I cannot aim at providing the complete picture, I do fill in a few more details, in order to have an assessable version of the view. By doing so, I conclude that a claim by Mark Schroeder is correct, which is that a proportionalist Humean theory of reasons can be summarized as the view that what one has most reason to do is whatever
would best promote her strongest desires. Not only do most people find the view so summarized unappealing, but when combined with contractualism, it raises the very difficult Problem of the Fool.

And so, in Chapter 5, I address this very old, very difficult problem, and claim to make some real progress. By expanding on a classic Hobbesian solution, I argue that I am able to satisfactorily answer the Fool and to solve another serious problem raised by the solution. Having claimed this success, then, I re-raise the issue of HCA’s status as a ‘backup’ theory, and ask what it would need to do in order to be considered, instead, a ‘primary’ candidate view. In Chapter 6, I argue that HCA would need to answer a wide swath of objections, which I then proceed to address. Although I certainly do not show that HCA is, after all, explanatorily perfect, I conclude in Chapter 7 that it is so successful that, given its virtue of modesty, it ought to be considered a live option when evaluating theories of moral reasons.

0.3 Beginning with Contractualism

An investigation of this size must begin somewhere, and it would be foolhardy to begin by trying to argue for the correct foundation for my view. So instead, I begin by adopting the moral contractualist framework put forward by T.M. Scanlon – not only because it seems to me to be a plausible moral theory, but for the following reasons as well.

First, as I will discuss in more detail in the Conclusion (and as I hope is apparent in Chapter 1), one of Scanlon’s great contributions to moral philosophy concerns his views on ‘the motivational basis of morality’. As a result, he asks exactly the question that I want to answer – what reason does one have to act morally – and provides, on his way to an answer, a wealth of insights concerning the plausibility of a variety of kinds of answers. Thus, although I will
eventually reject Scanlon’s own answer to the question ‘why be moral’, I take it that he has provided one of the most helpful starting points for one who wishes to answer the question. Of course, if one was genuinely skeptical of contractualist moral theories, this first reason would not exert much pull on her, as Scanlon’s account of the reason to be moral is clearly tied to his account of morality. So, for the sake of my overall philosophically modest methodology, I want to offer one more reason to take contractualism as a promising starting point.

In short, I take Scanlon’s contractualism as my point of departure because I contend that contractualism is plausibly true, regardless of whatever else is. That is: Scanlon’s (summary) criterion of permissibility is that permissible action is justifiable action; or, more carefully, that impermissible actions are those that cannot be justified to others on grounds they could accept. But, so understood, the view is overwhelmingly plausible, and no one has reason to reject the view.3 Let me explain what I mean.

Of course, there is a sense in which many philosophers reject contractualism: namely, they reject that impermissible action is impermissible because it is unjustifiable. However, the criterion of permissibility as I’ve stated it does not claim that some act’s being unjustifiable explains its status as impermissible; rather, it simply states that impermissible actions are those that cannot be justified, or in Scanlon’s language, an act is impermissible if and only if it could not be justified. But of course, that the truth conditions of an act’s being impermissible, and that same act’s being unjustifiable, are always the same, does not imply any particular explanatory or

3 Indeed, Scanlon notes that “[t]he idea that an act is right if and only if it can be justified to others is one that even a noncontractualist might accept” (1998, 189). Further, in a footnote to that observation, he cites Judith Jarvis Thompson’s suggestion that this biconditional “is arguably a necessary truth” (1990, 20 f. 15; see also 30 f. 19).
causal relationship between the two. For ease of discussion, we can distinguish between 

*substantive* contractualism – the view that the unjustifiability of an act *explains* its wrongness – from *formal* contractualism – or the view that a biconditional holds between an act’s being unjustifiable and its being impermissible. Formal contractualism, then, simply does not take a stand on the truth of substantive contractualism.

As a result, formal contractualism need not be rejected by anyone, almost regardless of her other philosophical commitments. The reason for this agreeability is that the biconditional between unjustifiable action and impermissible action allows that the explanation could go in the opposite direction of that assumed by substantive contractualism; that is, it could be the case that acts are unjustifiable *because* they are impermissible. And for anyone who believes that there are independent standards of permissibility, this should seem overwhelmingly plausible. If an act is impermissible as a result of it failing a test of universalizability, or because it brings about worse consequences than another action, or for any other independent reason, it seems that its wrongness ought to count against the action sufficiently that it renders the act unjustifiable. No matter what reasons I offer in favor of the act I am considering, if it is wrong, then presumably (on most views) those reasons will fail to justify the act. But if this is the case, then most anyone ought to accept formal contractualism. Whether the direction of explanation is from justifiability to permissibility, or from permissibility to justifiability, the two seem to co-travel; and that is all that formal contractualism requires.

Scanlon argues for the truth of substantive contractualism, and I, also, believe that substantive contractualism is an independently plausible moral theory. However, a methodologically modest reason for beginning my investigation with contractualism is that the
view I develop – HCA – can be assessed as a view employing mere *formal* contractualism. One need not accept substantive contractualism in order to accept the view that moral action is justifiable action, that reasons are explained in a Humean way, and that the wide-spread desire for intimacy generates a reason for many people to act justifiably. And while restricting oneself to a reading of contractualism as merely formal would likely rob some of my arguments of their force (see especially Chapters 6 and 7), it is still a virtue of HCA, I believe, that it can be embraced by those who reject substantive contractualism.

There are clear reasons, *for me* in particular, then, to begin this investigation with contractualism, as I believe it is likely true, and Scanlon has provided a rich exploration of the relationship between contractualism and the reason to be moral with which to begin. However, there is the additional, modest reason that much of this dissertation can be read, without loss, as an investigation concerning formal contractualism. Doing so does not undermine the ability of HCA to explain the reason to be moral, as it simply holds that the desire for intimacy explains one’s reason to act justifiably. But if justifiable action is permissible action, it doesn’t matter which direction the explanation runs: by explaining one’s reason to act justifiably, HCA explains one’s reason to act morally.

0.4 Fragmentation of the Moral

Before diving into the meat of the dissertation, there is one last ground-clearing point that it will be helpful to make. In *What We Owe to Each Other*, Scanlon claims that he is, with his criterion of wrongness, giving an account of only part of what is often referred to with the language of morality – namely, the part concerned with, well, what we owe to each other. He believes that he can work in this seemingly piecemeal way because of what he calls the
‘fragmentation of the moral’. On his view, “most of us commonly use the terms ‘moral’ and ‘morality’ to refer to a diverse set of values, and…while contractualism characterizes a central part of the territory called morality, it does not include everything to which that term is properly applied” (1998, 173). Some people think that there are genuinely moral issues concerning sex, for instance, that do not have to do with what we owe to one another, and if this were right, then these aspects of sexual ethics would need a different explanation from the one Scanlon gives.\(^4\) Or, less controversially: animal ethics and environmental ethics seem to concern ‘moral’ claims, but are not necessarily about what we owe to each other.\(^5\)

The first point that I want to make is simply that I take Scanlon’s move to a more restricted domain to be well-motivated, and so I follow suit.\(^6\) It does, indeed, seem to me that we refer to many different sorts of domains with the language of morality, and it would in fact be rather surprising if all of them could be explained in the same way. Further, it does not seem arbitrary to focus on what we owe to each other, as this seems to be a central, and centrally important, domain of morality. If it is wrong to eat non-human animals, to chop down an ancient tree to improve one’s view, to fail to develop one’s talents or to engage in some particular sexual

\(^4\) Obviously, there are genuinely moral issues regarding sex that are also about what we owe to each other: it is impermissible to force one to have sex, for instance, but that is not a claim in ‘sexual ethics’ distinct from what we owe to each other. The issue here, rather, is whether there might be true moral claims regarding sex that do not invoke the standards of what we owe to each other – perhaps regarding masturbation, homosexuality, promiscuity and the like.

\(^5\) Note the ‘necessarily’ modifier, as surely some claims about what we owe to each other will concern animals and the environment. For instance, you have a reason not to hurt my dog as a result of what you owe me. And we all have a reason to slow our carbon production as a result of what we owe one another. However, a concern with some animal’s pain, or for the majesty of the great, old Redwood, does not seem to be a concern with what we owe to each other.

\(^6\) I also follow suit in continuing to speak in terms of ‘morality’, ‘moral facts’, ‘moral considerations’, etc, though I intend ‘interpersonal morality’ or ‘what we owe to each other’. 12
lifestyle, the seriousness of these wrongs does not seem to approach the seriousness of violating what we owe to each other. Which brings me to my second point.

Many people seem to think that the restriction of the scope of morality by contractualism is a bad-making feature of the view. It is said that contractualism has a ‘problem’ accounting for animals, say, or the environment. However, what I want to suggest is that a view like the one I will develop here – that is, like HCA – is particularly well-suited to explain the wrongness, and the seriousness of the wrongness, of varying domains. As will likely become clear in Chapters 1 and 7, I think that Scanlon’s claims concerning the importance and priority of justifiability are some of the most insightful aspects of his view. We care about what we owe to each other as much as we do because it concerns how we relate to others, and how they see us. Violating the standards of what we owe to each other is not minor, because even if the stakes are relatively low (no one was harmed, reparations are easy, etc), intentional acts of wrongdoing reveal about someone that they do not accord us the standing to partially determine the grounds on which we live together. A commitment to acting rightly is a commitment to valuing other agents as the sorts of creatures to whom justification is owed.

The view offered in this dissertation – HCA – claims that our reasons to act are grounded in desires, and so is well-suited to explain both the diversity of concerns that demarcate the moral domains, as well as the varying levels of seriousness of the different domains. To take an illustrative example: I will argue, in Chapter 2, that HCA can explain a wide-spread reason to be moral by reference to the desires to live justifiably, and to have particular, intimate relationships. However, these are not the only desires that agents often have that strike us as morally relevant. We often have sympathetic desires, or desires to prevent suffering, and as the case of animal
ethics shows, these desires do not generate reasons concerning only how to interact with other people. Our sympathetic desires also provide reason not to cause gratuitous harm to other sentient animals, and this desire may well be what grounds our reason to treat animals well (that is: such a desire might ground ‘the moral reason’ in the context of animal ethics).\(^7\)

I won’t here take the time to investigate how revisionary an ethic such a desire would entail, or whether it should be viewed as overly-strong or overly-weak. But what seems plausible regardless of how the view works out, is that the domain of what we owe to each other is grounded in a fundamentally different way from the domain of animal ethics. Further, then, they are grounded differently in a way that allows us to explain the difference in how important we take them to be. Even the most dedicated vegans do not (typically!) treat their omnivorous colleagues as if they were mass-murderers, and there is a good explanation for this: our reasoning concerning animals does not reveal what standing we grant other people. While there are some ways in which, by sharing a planet, talk of animals as resources or parts of an ecosystem will require interpersonal justification, discussion of much of animal ethics does not require that we take one another to have a particular kind of value. And so, much as we may be concerned with animals, our concern is fundamentally different from the concern that others act justifiably, and it is revealed in the difference in seriousness with which we treat the domains of animal ethics and what we owe to each other.

\(^7\) Interestingly: HCA might then vindicate the view that Robert Nozick calls ‘Kantianism for people, Utilitarianism for animals’ (1974, 35-42), and without the distinction being problematically \textit{ad hoc}. It just turns out that our moral reasons differ in the realms of animal ethics and what we owe to each other. Nozick himself thought so little of utilitarianism that he found it inappropriate even for animals; but I, myself, find this basic way of grounding the two moral domains quite promising.
0.5 Conclusion

I now turn to the first step in the project. As I previewed above, there are good reasons to begin with Scanlon’s contractualism for any investigation concerning the motivational basis of morality. In short, his thoughts on the subject are nuanced and insightful, and the framework that he gives for investigating the reason to be moral – as well as the criteria that he gives for what would count as a successful view – deserve close attention. In what follows, then, I hope to begin the long project of developing HCA, but while, at the same time, devoting considerable care and attention to the independently interesting project of engaging with Scanlon’s view.
CHAPTER 1

A complete moral theory must provide an account of what is often called ‘moral motivation’, the job of which is to explain ‘what reason one has to be moral’. This much tends to be agreed upon. Less clear, however, are the details concerning what is being demanded. The term ‘moral motivation’ seems to imply that what is sought is primarily a psychological account, concerning, perhaps, what does or could motivate an agent to moral action. The language of ‘the reason to be moral’, on the other hand, has a normative ring to it: the explanation being sought here concerns what reasons there are. Thus while there is largely an agreement on the need to provide some kind of account concerning the link between morality and motivation, the details are unclear in the abstract, and anyone hoping to answer the call must do some clarificatory work at the beginning.

T. M. Scanlon, in making a case for his unique brand of contractualism, does just this. In his words, a satisfactory moral theory must “explain the reason-giving and motivating force of judgments of right and wrong” (1998, 147) – a challenge that clearly has both motivating and normative components. Rather than explaining ‘why one ought to be moral’, or ‘why one has reason to be moral’, Scanlon interprets the moral motivational question as asking ‘how the fact that an act is wrong provides a reason not to do it’ (Scanlon 1998, 147-149).1 And this question

---

1 Although it is convenient to speak of actions, the moral motivational reason might also apply to ways of reasoning as well, and so may rule out acting directly on the basis of its recommendation. On Scanlon’s view, for instance, paradigmatically moral actors (as we will see below) act for the sake of living with others on terms that they could accept. However, those terms may require that one not reflect on what is acceptable before considering every action or inaction. Thus, Scanlon thinks that one could reasonably object to a set of principles that allowed one to deliberate about whether to, say, torture for fun, or to rape. In this way, he gets around the ‘one thought too many objection’ by appealing to a moral motivational reason that
can be read in two different ways. On the one hand, we want to know the answer to an empirical question, namely: what those who care about morality are moved by; and on the other hand, we are demanding a kind of justification: what is it about morality such that everyone has a reason to care about it? ‘The’ question of moral motivation, then, is actually two questions. For this reason, Scanlon suggests that a better name for the challenge of ‘moral motivation’ is ‘the motivational basis of morality’, as this name wears its dual character more clearly on its face.\(^2\)

In *What We Owe to Each Other*, Scanlon outlines the central problem facing any account of the motivational basis of morality, providing his own, contractualist solution. The problem, Scanlon thinks, is to navigate ‘Prichard’s Dilemma’ (PD), which arises from the demand that explanations of the reason to act morally be both (a) helpfully explanatory, and (b) relevant to morality. According to PD, all moral theories must explain the reason to be moral by reference either to a moral consideration, or a non-moral consideration. Explanations by reference to moral considerations, however, are trivial and unhelpful (thereby violating (a)), while explanations by reference to non-moral considerations offer implausibly external incentives to be moral (thereby violating (b)). Scanlon’s solution to this dilemma is to explain the reason to be moral by reference to a moral-but-still-helpful value – namely, the value of living with others on terms that all can accept.

In this chapter, I begin the long-term project of developing my own account of the reason to be moral by engaging with Scanlon’s careful and influential view of the motivational basis of moral theories. Although Scanlon notes this complication only in passing, it will come to play a large role in the concluding sections of this chapter.\(^2\) He does, however, continue to use the exceedingly handy phrase ‘moral motivation’, and so I will do the same.
morality. And while I will eventually modify his account with a heavier hand, I will begin by accepting as many of Scanlon’s philosophical commitments as I can. In particular, I will not question the contractualist criterion of wrongness, the above setup of the moral motivational problem, or the account of reasons and value laid out in *What We Owe to Each Other*. My first goal, then, will be to argue that even given Scanlon’s commitments, there is at least one solution to PD in addition to the one he singles out. Second, I will argue that Scanlon’s own solution to PD suffers from some of the problems of the ‘unhelpful’ horn of the dilemma, while the solution I offer suffers from some of the problems of the ‘irrelevant’ horn. In response to the failure of these – what I take to be quite promising – candidate solutions, I then suggest a different strategy for solving PD: combining multiple solutions. If explanatorily helpful solutions tend to be irrelevant to morality, and solutions that are relevant to morality tend to be unhelpful, then a hopeful model for solving the problem is to offer multiple explanations. Of course, this cannot be done in an *ad hoc* way, and so in the final sections of the paper, I demonstrate how the individual solution that I offer and the one that Scanlon offers can be naturally combined to form part of the same explanatory story.

1.1 Prichard’s Dilemma

According to Scanlon, any attempt to provide an account of the motivational basis of morality faces a difficult challenge, which he calls ‘Prichard’s Dilemma’, so-called after H.A. Prichard’s description of a similar dilemma in his essay, “Does Moral Philosophy Rest on a Mistake?” (1949). The challenge here is that is that the question, ‘why be moral’ seems to require, on the one hand, a moral answer. One has reason not to act immorally because the fact of an action’s being wrong is a reason not to do it. But this, of course, is not much of an answer
at all; it takes the reason-giving force of morality for granted, when the reason-giving force of morality is precisely what we want explained. The same is true of other, closely-related answers to the moral motivational question. In a particularly Kantian moment, one might be tempted to say that one has reason to be moral ‘because it is one’s duty’, but this answer fares no better. It simply assumes that duties have reason-giving force, which is to say that morality has reason-giving force. On this horn of the dilemma, the answers are what Scanlon calls ‘trivial’, or what I call ‘unhelpful’, as the explanations take for granted precisely the phenomenon that we are trying to explain.

On the other hand, one might attempt to provide a clearly non-trivial, obviously helpful explanation of morality by reference to a non-moral consideration. For example, one might think that the most satisfying way to account for the reason to be moral would be to show that being moral would make one happy, or would otherwise be in one’s interest. But on this horn, we face a different challenge, as such an answer seems to provide one with the wrong kind of reason to be moral. If I ask why I have reason to donate to famine-relief, there seems to be something inappropriate about citing how good it will make me feel in response. To do so, Scanlon says,

\[\text{\footnotesize{\begin{enumerate}}\]

\[\text{\footnotesize{\item Whether this was in fact believed by Kant, it is a popular interpretation of claims he makes in the \textit{Groundwork} (1996, 37-109) and elsewhere. Scanlon, himself, seems to think that Kant holds such a view (1998, 148-149).}}\]

\[\text{\footnotesize{\item The popularity of this move can be seen as far back (at least) as Plato’s \textit{Republic} (2004), in which Socrates accepts Glaucon’s challenge to demonstrate how the just person is better off than the unjust, regardless of how the world might conspire against the just person. So desperate was Glaucon and his fellow interlocutors for an account of a self-interested reason to be moral that they demanded of Socrates an explanation of how it is in one’s interest to have justice ‘only in the soul’, such that one is better off being just even should doing so make him worse off by all external measures. Plato’s Socrates spends the entirety of the \textit{Republic} attempting to discharge this burden.}}\]

\[\text{\footnotesize{\end{enumerate}}\]
would be to offer an ‘implausibly external incentive’ for acting rightly. And the whole Kantian tradition seems to recognize this danger, as it tells us that such an act would lack true moral worth; what one *ought* to do is give to famine relief out of respect for the moral law. While considerations of one’s happiness might in fact motivate some people to do some right actions, such considerations are not what we would expect the paradigmatically moral person to cite as her reason for acting. According to this set of intuitions, then, an account of the motivational basis of morality must also be relevant to morality.

The challenge of PD, then, demands of any account of moral motivation that it be both helpfully explanatory and relevant to morality. While Scanlon makes this challenge fairly quickly, I take it that his intuitions are widely shared. Accounts that explain our reason to be moral in terms of our happiness or interests do, in fact, seem to have a problem explaining the special value of acting out of duty, while those that explain our reasons by reference to very closely-related concepts like ‘duty’ are unsatisfying.

Although Scanlon’s analysis of the two, competing *desiderata* for a theory of moral motivation is compelling, I think (and I think that Scanlon knows) that the challenge isn’t a true dilemma. Instead, candidate moral motivational accounts seem to be able to succeed more or less with regards to the criteria of relevance and helpfulness. This can be seen even in the two ‘trivial’ candidates, as the Kantian view that one has a reason to act rightly because it is one’s duty is slightly less trivial (and therefore more explanatorily helpful) than the view that one has a reason not to act wrongly because the fact of an action’s being wrong just is a reason. Similarly, although many bristle at the idea that one has a reason to act rightly because it would make her happy, it is plausible that this is a more morally relevant answer than ‘one has a reason to act
rightly because it would result in more facial muscles being used’ (supposing that it were true).
And so it in fact looks like PD is not a genuine dilemma, but a challenge of meeting two,
seemingly contrasting desiderata, solutions to which satisfy these criteria to more or less a
degree. Scanlon says similarly,

Answers [to PD] can thus be arrayed along one dimension according to their evident
moral content, ranging from those that appeal to what seem most obviously to be moral
considerations (thus running the risk of triviality) to those having the least connection
with moral notions (thus running the risk of seeming to offer implausibly external
incentives for being moral)” (1998, 150).

My final commentary on this point is that Scanlon’s set-up of the moral motivational
problem helps us to see why PD, although not a true dilemma, is so challenging. The moral
motivational question, remember, has two parts: a successful theory must explain both what
those who care about morality are moved by when they act rightly, as well as why everyone has
reason to care about doing the same. When we focus on the latter half of the challenge, we are
pushed in the direction of looking for incentives: what kind of reason can we come up with that
grips everyone, or nearly everyone? Such non-moral incentives, however, remind us of the
former half of the challenge, and we find ourselves dissatisfied with a view that has
paradigmatically moral actors acting for non-moral reasons. Thus, we simultaneously want the
answer to have moral content, to satisfy the former half of the challenge, while being devoid of
moral content in order to satisfy the latter. A satisfying solution to Prichard’s Dilemma must
somehow convince us either that a moral explanation can be helpful and non-trivial, or that a
non-moral explanation can still be morally relevant.\textsuperscript{5} On my understanding of Scanlon’s account, he opts for the former strategy, to which I now turn.

1.2 Scanlon’s Account: The Good of Mutual Recognition

According to Scanlon’s well-known criterion of wrongness, an act is wrong just in case it would be disallowed by any set of principles that others, similarly motivated, could not reasonably reject (1998, 4; 153).\textsuperscript{6} Or, in Scanlon’s own shorthand: an act is wrong if it could not be justified to others (where ‘able to be justified’ just means allowed by rules that others could not reasonably reject). The challenge of accounting for the motivational basis of morality, then, is to explain why the fact that some act is unjustifiable provides a reason not to do it; and the additional challenge of PD demands that Scanlon’s explanation be both helpful and relevant to the content of morality. Scanlon believes he has such an explanation in the value of mutual recognition.

‘Mutual Recognition’ is Scanlon’s name for a particular relationship in which people are capable of standing – in particular, of the relationship that one has with others when both she and they are living on terms that no one could reasonably reject. While Scanlon admits that standing in mutual recognition with others is much less personal than relationships like friendship (1998, 162), the description of this way of living with others as a relationship seems to him to be phenomenologically accurate. What happens when I realize that an action of mine is

\textsuperscript{5} Or, it must reject the implicit assumption that there is a single reason to be moral. This is the route that I will take in the final sections of this paper.

\textsuperscript{6} Recall from the Introduction that this is actually Scanlon’s account of interpersonal morality, or what we owe to each other. Recall also that I am simply following his lead here – referring only to this more restricted domain, but continuing to use the language of ‘morality’ and ‘wrongness’ to refer to it.
unjustifiable, Scanlon thinks, is that I sense a shift in the way I am relating to others; I am no longer living with those around me on terms that they could reasonably accept (whether they do, in fact, accept them or not). This feeling of estrangement one gets from impermissible action, as well as the positive ‘pull’ one feels from moral action, stem from the “positive value of living with others on terms that they could not reasonably reject” (1998, 162).

So how, exactly, does the relationship of mutual recognition constitute the kind of explanation sought here? Although Scanlon says much about the issue, he never addresses the question in a direct way, and so providing an answer on his behalf requires some interpretation. What he does tell us is that standing in the relation of mutual recognition to others is ‘appealing in itself’, and ‘worth seeking for its own sake’ (1998, 162). But standing in such a relationship, the very content of which is living with others on terms they could accept – requires actually acting according to terms that they could accept. This is a constitutive requirement, as the relation of mutual recognition is simply constituted by living with others according to the contractualist’s criterion of permissibility. Scanlon gets close to saying this, claiming that for the moral person, moral requirements “are not just formal imperatives; they are aspects of the positive value of a way of living with others” (1998, 162).

---

7 One might question this necessity. In the same way that one can be in love without, at some particular moment, acting lovingly, it may be that one can stand in mutual recognition of others and still occasionally act unjustifiably. On my understanding of Scanlon’s view, this concern would cause him problems; and indeed, given that I will eventually adopt a view sufficiently similar to Scanlon’s, it will eventually cause me problems. However, I am here merely trying to understand Scanlon’s view, not to criticize it, and I do, in fact, think that a constitutive requirement makes the most sense of his solution to the problem of moral motivation. My own solution to the version of this concern that I inherit will be addressed in Chapter 5.
The relation of mutual recognition thus provides a particularly tight explanation of why an act’s wrongness provides a reason not to do a thing. I call this form of explanation *constitutive*, as the relation between the act’s wrongness and the reason not to do it is that the former partially constitutes the latter. One ought not, for instance, cause gratuitous harm, because others could reasonably object to our so acting – that is, we could not justify our action to others by appeal to principles that we all accept. And why is that a reason not to do it? Because acting justifiably *constitutes* ‘an aspect of the positive value of a way of living with others’ – because living with others on *permissible* grounds is appealing in itself.

### 1.3 Promise and Peril of Scanlon’s Account

A key virtue of the above moral motivational account is, Scanlon thinks, that it offers a satisfying solution to Prichard’s Dilemma. This is because the account explains the reason to be moral by reference to an “ideal of relations with others which is clearly connected with the content of morality and, at the same time, has strong appeal when viewed apart from moral requirements” (1998, 155). Given Scanlon’s criterion of permissibility, the relation of mutual recognition just is living with others *permissibly*, or perhaps living with others on *permissible grounds*. The moral motivational account thus has clear moral relevance. However, Scanlon thinks that it is not thereby trivial or unhelpful, because such relationships are also ‘appealing’, or ‘worth seeking for their own sake’. By explaining the reason one has to act permissibly by reference to mutual recognition Scanlon takes himself to be saying something genuinely enlightening. That some act is wrong provides one with a reason not to do it because acting wrongly would do violence to the relation of mutual recognition. The good of standing in such
relations accounts for the normative pull we feel when confronted with moral decisions, as acting in a way that removes us from, or does damage to, such relationships would constitute a loss.

There is much to like about this account. While I will, in short order, question whether it really does navigate between the Scylla of unhelpfulness and the Charybdis of relevance, it at least provides a model of how one could do so by utilizing a form of relationship as the way to explain the reason-giving force of morality. As a contractualist, Scanlon holds that the content of morality (remembering that for him, the relevant domain is ‘what we owe to each other’) concerns what people could reasonably agree to. Thus, moral life essentially involves getting along with others in some way, and the very attractiveness of contractualism (for those of us who find such views attractive) is due to this relational feature. So the idea that what explains the normative ‘pull’ of moral considerations must have to do with our relation to others is a natural one, befitting of a contractualist theory.

My suggestion, then, is that Scanlon’s major success concerning his account of the motivational basis of morality is in pointing to human relationships as the proper explanation of why moral considerations provide reasons for action. For a theory that makes moral facts depend on how people relate to one another, this seems like a promising move. If we add to this success Scanlon’s suggestion that there is a real phenomenological plausibility here, then the case gets stronger. If, finally then, it were the case that the particular relationship Scanlon singles out as the relevant explanation did, in fact, seem to solve Prichard’s Dilemma, then Scanlon’s position here would look like an unmitigated success. However, I want to suggest that his position is rather a mitigated success. While I think that Scanlon’s solution is promising, it does not succeed as well as he hopes.
The promise of Scanlon’s attempted solution to PD is in the idea of explanation by reference to relationships. Moral action is not, itself, a relationship, and so the *explanans* here adds something genuinely new. And, if my observations above are on target, then what is added to the explanation here is of the right kind: since contractualism takes the morality of an act to depend on whether the actor could get along with others in a particular way, explanation of morality’s force by reference to relationships is a promising way to go. However, the particular relationship of mutual recognition is characterized by living with others according to the contractualist criterion of permissibility, and so we might in fact call this the ‘moral relationship’. And the presence and role of such a moral relationship in Scanlon’s theory is unsurprising given his Kantian leanings. Standing in the relation of mutual recognition to all others is reminiscent of living in the kingdom of ends. As Scanlon explains in his chapter on value, the proper way to value persons is not to promote them, but rather to see them as beings to whom justification is owed; and this is, according to Scanlon, how one ‘respects another’s rational nature’ (1998, 106). Living with others on terms they can accept, then, is valuable in a way that is similar to the way in which living in the kingdom of ends would be valuable: to do so would be to live with others in such a way that all are treated as their value requires.

---

8 Although Scanlon resists the association with Kantianism at times (see his “How I’m Not a Kantian” in Derek Parfit’s *On What Matters* (2011)), he also acknowledges his roots. For instance, in the introduction of *What We Owe to Each Other*, Scanlon admits the similarities between his view and those that inspired and were inspired by Kant, such as Jean-Jacques Rousseau (1988) and John Rawls (1971). This helps to distinguish him from other contract theorists, such as Thomas Hobbes (1968), and from those today who take their lineage more from Hobbes than from Kant, such as David Gauthier (1986).
Drawing out the Kantian sympathy here reveals, I think, the danger in Scanlon’s account: there is too much moral content in Scanlon’s explanation for it to successfully navigate PD. His solution, although more satisfying than the most radical positions on the unhelpful horn of the dilemma, defines moral action as justifiable action, and then explains the reason to act morally by reference to the value of living with others on justifiable terms. If one is not already gripped by the value of acting justifiably, it is unclear why the introduction of living with others on justifiable grounds should help to make this clear; it is no more obvious why one should care about living with others in mutual recognition than about acting justifiably. Thus, while Scanlon’s solution – like the Kantian one he criticizes – clearly succeeds in showing why those who care about morality have reason to act morally, it is less successful in helping us to understand why acting morally is something that everyone has reason to care about.

This is not to say that Scanlon has made no progress. First, it should be noted that just by giving a criterion of permissibility and then attempting to explain the reason to act according to such a criterion, Scanlon has already made a small step in providing a less trivial answer. Whereas the most egregious failures of triviality either simply insist that morality is reason-giving, or explain morality’s reason-giving force by reference to other, equally moral and closely related concepts like ‘duty’, Scanlon’s account tells us what morality is. So the explanation of the reason to be moral isn’t simply in terms of what I called ‘the moral relationship’; rather, the explanation of the reason to be moral is in terms of living with others justifiably, when moral action is understood as justifiable action. While the difference here is nuanced, it seems clear that Scanlon has succeeded in providing a more explanatorily helpful answer than those occupying the pole at the ‘trivial’ end of the spectrum.
In addition, I stated at the outset that I think Scanlon has provided a model of the right way to answer Prichard’s challenge. By introducing the idea that the best solution will involve human relationships, Scanlon has moved the conversation forward another small step. Thus the ‘mutual recognition’ view of moral motivation is an improvement over the most trivial explanations in at least two ways: it provides an account of morality, and it picks out interpersonal relationships as a plausible candidate for the explanans of a contractualist moral motivational account. While I have charged that Scanlon’s view is ‘too close’ to the trivial end of PD, it doesn’t fail as egregiously as, say, the Kantian view. It is, however, still the same kind of unsatisfying as the Kantian solution; it is just unsatisfying to a lesser extent.

What this mitigated success sets up is the idea that there may be multiple solutions to PD that succeed to varying degrees at being relevant to morality while still being helpful. And indeed, this is precisely what I think is the case. In the remainder of this chapter, I will argue that Scanlon’s own arguments regarding the interaction between morality and relationships make possible another solution to PD – one which provides a clearly non-trivial explanation. This second solution, however, has problems that mirror the worries that Scanlon’s original solution faces, as it seems too close to the ‘irrelevant’ horn of PD. In the final sections, I employ both of these solutions (the one that Scanlon identifies and the one that I do) as evidence for the claim that, although PD may not be a real dilemma, it does challenge that no single consideration is likely to generate an explanatory story that is completely satisfying, as all such accounts are likely to seem, to some extent, either morally irrelevant or trivial. I then solve this problem by adopting both explanations discussed – the one raised by Scanlon and the one raised by me – as part of a single explanatory story.
1.4 Friendship and Justifiability

Scanlon holds that a full explanation of the motivational basis of morality must be able to account for the importance we place on others responding to moral reasons, as well as the priority that such reasons seem to have in our moral deliberation (1998, 158-168). It is in his discussion of this second requirement – that of solving the problem of priority – that Scanlon makes the argument on which the current section will focus.

Explaining the priority of morality’s reasons over other, non-moral reasons is a special challenge, Scanlon thinks, because it threatens to prove too much; put briefly, it threatens to show that persons must hold all other values loosely, since any project- or value-based reason could be overridden by moral considerations. Scanlon’s proposed solution to this problem is to show that central cases of seemingly non-moral projects do not challenge the priority of moral reasons. His test case in this argument concerns the good of friendship. According to Scanlon, then, the good of friendship is not threatened by morality’s priority because “[f]riendship…involves recognizing the friend as a separate person with moral standing – as someone to whom justification is owed in his or her own right, not merely in virtue of being a friend” (1998, 164). Friendship thus presupposes, or depends on, moral recognition. There is no conflict between morality and the good of friendship, because friendship requires that we recognize a friend’s moral standing.

We might summarize this point as follows: the good of friendship depends on the good of mutual recognition, as having the specific type of friendship relation that Scanlon is considering requires seeing one’s friend first as someone to whom justification is owed. In order to motivate this strong claim, Scanlon gives the example of a ‘friend’ who one day confides in you that he
would steal a kidney on your behalf, should the need arise. Having such a friend would be ‘unnerving’, Scanlon says, “because of what it implies about the “friend’s” view of your right to your own body parts: he wouldn’t steal them, but that is only because he happens to like you” (1998, 164-165). The power of this example, I take it, is due to the discomfort at the thought of letting someone so unprincipled close to us. Absolutely, we want friends to act for special-to-us ‘friend-reasons’ – because they like us – but we also want assurance that not all of the ways in which they treat us are due to these idiosyncratic, warm feelings. In order to have a genuine friendship with someone, Scanlon suggests, both the potential friend and I must be committed to a general relation of mutual recognition with others. Only when both parties have the kind of moral assurance thus given does it become possible to relate to one another as friends. There is thus no incompatibility between the good of friendship and that of morality.

Scanlon has, with this argument, proven much more than the compatibility of friendship with morality, however. He has actually proven that friendship depends on the moral relationship. It is thus not simply the case that friendship will not pose a challenge to morality; but friendship will not pose a challenge to morality because friendship requires morality. This is a very strong claim, and there are reasons to be skeptical of it. Scanlon admits, for instance, that there is no reason that one cannot have a variety of relationships with others of a kind that we often call friendship even if one or both parties are uninterested in living with others on terms they can accept. However, Scanlon holds that there is some particular, intimate and valuable relationship (which may not map on perfectly to the way we use the word ‘friendship’) that
works as he suggests. It both presupposes moral concern on the part of both parties, and it is a relationship worth seeking for its own sake (1998, 165). And, strong claim or not, something like this doesn’t seem radically implausible. Indeed, such a position has quite the pedigree, going back at least to Aristotle, who held that only the truly virtuous were capable of the highest level of friendship (1999, Books VIII-IX).

However, it is not yet my goal to argue for plausibility. Instead, my goal for now is to point out that, given Scanlon’s commitments expressed in this discussion of friendship, he has available another explanation of the motivational basis of morality. In the remaining sections of this chapter, my plan is to draw out what the explanation is and how it works, after which I will use the seeming weaknesses of both it and Scanlon’s own account to argue for a ‘complex’ model of moral motivation.

1.5 The Friendship-Based Account of Moral Motivation

---

9 It may seem that Scanlon has here highly qualified his success on solving the priority challenge. It is only this close, intimate friendship that he has shown does not challenge morality’s priority. However, there are two things to say in response. First, a very close friendship is likely the kind that ever threatened morality in the first place; one does not typically worry, I take it, that perhaps I ought to do terrible things for a mere acquaintance. It is the intimacy and meaningfulness of the friendship relationship that both seems to raise the problem of priority in the first place, and which seems to solve it. Second, Scanlon still admits that not all ‘genuine friendships’ throughout time and culture will plausibly fit this model. So, for instance, he raises Achilles and Patroclus as candidates for having a friendship that does not have the character he has described – a kind of friendship that does, in fact, challenge the priority of morality. In response, Scanlon says quickly that “the claim that [morality] would clash with the demands of this idea of friendship is a much less forceful objection to morality as I describe it than the charges, to which I have responded, that it is incompatible with friendship as we understand it, or with the conceptions of friendship that we have most reason to value” (1998, 165).
Scanlon’s account of the motivational basis of morality is that the good of mutual recognition explains one’s reason to act justifiably. Such an account is intended to be helpfully explanatory because the *explanans* – the relation of mutual recognition – provides us with a recognizably appealing relationship that requires (constitutively, I’ve suggested) justifiable action, which on his contractualist view just is moral action. Scanlon’s own account has thus given us a model for how alternative accounts may be generated: we are to look for non-moral goods (although, when we go to check our success against PD later, we will find that we want it to be a non-moral good that somehow seems relevant to morality) the having of which requires that one act justifiably. My current suggestion is that Scanlon’s own argument concerning the relationship between friendship and a requirement of justifiability provides such an account.

The kidney-thief example from above is supposed to make plausible the claim that standing in the relation of (some kind of) friendship requires acting justifiably. This is because we are ‘unnerved’ by becoming close to someone who is uncommitted to justifiability, and so we shut such people out of our more intimate relationships. More carefully: the example suggests that the having of such friendships requires relating to others as ‘mutual recognizers’, and relating to others as a mutual recognizer requires, according to Scanlon, acting justifiably.

This argument concerning the role of friendship thus makes possible a further explanation of one’s reason to be moral. We all have reason not to steal a kidney (even for a friend), because kidney thieves cannot live with one another on terms that all can accept. However, I suggested at the end of section 1.3 that this explanation may have too much moral content to be genuinely helpful as an explanation. We might reformulate this complaint as a demand for a *further* explanation; that is, *why* does the fact that kidney thieves cannot live with one another in mutual
recognition give one a reason not to be a kidney thief? And now we have an answer: because failing to be a mutual recognizer blocks one from the ability to have friendships. Thus, leaving out the middle term in the explanatory story,\textsuperscript{10} we can summarize as follows: according to Scanlon’s set of arguments here investigated, one has reason to act justifiably because failing to do so blocks one from the good of friendships.

1.6 Intimacy and Vulnerability

In the brief explanatory story above, I have tried, despite adding very little to what he explicitly states, to do something new and different with Scanlon’s arguments. However, this has left some matters vague. The goal of the following two sections is to step back from my agnosticism concerning Scanlon’s argument, and to do my best to turn the considerations above into something closer to a well worked-out position, while taking care not to import any assumptions that Scanlon would reject.

Scanlon claims that having access to the goods of friendship requires relating to others as a mutual recognizer, but he leaves it unclear why this is the case. Our intuitions about the kidney-thief case are supposed to convince us that it’s true, but we are offered no explanation of

\textsuperscript{10} That the middle term is part of the story will, however, be an important part of my final solution. I drop it for now only because the demand for an explanation of one’s reason to be moral, like all demands for explanation, seems to me to be asking for the \textit{ultimate} explanation, and so the summary seems warranted. For example: If my friend Pembe asks me why Marc Marquez crashed his motorcycle in the MotoGP race, I could reply, ‘because he flew off the motorcycle’. However true, this is not the final explanation that my interlocutor was looking for. So she may ask me again: but why did he fly off? Now, when I reply, ‘well, because he stabbed the front brakes too suddenly while leaned over, which upset the suspension’, I have finally seemed to discharge the explanatory burden. When Pembe relates the explanation to others, she is now likely to leave out the middle term in the explanatory story. Marquez crashed because he stabbed the brakes while leaned over. This summary of the ultimate explanation is what Pembe (and likely others) are looking for.
the property or properties of friendship that are responsible for this connection with morality. We thus do not know whether friendship is unique in its moral requirement, or whether there will be a principled way of determining a set of relationships that require a foundation of mutual recognition. My suggestion is that the latter is the case, and that the property responsible for friendship’s connection to morality is \textit{intimacy}; on this hypothesis, the more accurate conclusion to draw from Scanlon’s insight is that \textit{intimate relationships} require a foundation of mutual recognition.

My reasoning for this move is as follows: Scanlon emphasizes that it would be ‘unnerving’ to be friends with someone who would steal a kidney. However, I want to know what it is about having such a friend that is unnerving, to which Scanlon replies that such a friend sees no deep, principled reason not to take \textit{your} kidney – his aversion to this is based merely on his contingent, warm feelings. But this doesn’t seem to be a complete explanation either. So what is it about acting only on one’s warm feelings that should make us uncomfortable? My interpretation of Scanlon’s insight is that we want a particular kind of assurance of good will from those we choose to let close to us. The unease that comes from having a friend who is a kidney-thief is due to his physical and emotional proximity, plus the knowledge that he doesn’t see anything wrong with taking another’s organs (so long as the target is not someone he likes). But this aspect of friendship – the vulnerability of allowing another to be close to one – does not seem to be unique to friendship; rather, it looks to be a general property of any intimate relationship. By becoming intimate with another, we allow that person physical and emotional access to us, and this entails vulnerability; those with whom we are intimate are uniquely well-positioned to hurt us. Not only do we grant intimates physical
proximity, but we tell them secrets, share our feelings, and do a multitude of other things that open us up to being harmed. Scanlon’s insight thus makes good sense: we all have excellent reason – to which we reliably respond, he seems to believe – to form and maintain intimate relationships only with those who are mutual recognizers. This makes it the case that being a mutual recognizer is required in order to access the good of intimacy. And so we can explain the value of mutual recognition further in terms of the value of intimate relationships.

1.7 Explaining Reasons

While I think the story told above has a lot going for it, details have been smoothed over, both in Scanlon’s very quick telling of morality and friendship’s compatibility, and in my attempt to work up such observations into an argument. In particular, we might wonder whether it is really plausible that one is required to be a mutual recognizer in order to have intimate relationships – or put another way, whether it could really be true that it is impossible to have intimate relationships if one is not a mutual recognizer. Scanlon’s answer to the compatibility challenge suggests that he thinks so, but this is an implausibly strong claim, and so I do not want

---

11 The use of this universal quantifier seems as good a place as any to flag a natural worry, which I call the challenge of ‘Loyalty-Clan Counterexamples’ (thanks to Justin Weinberg for the name). The challenge goes like this: on my Scanlon-inspired view, having intimate relationships requires being a mutual recognizer. And perhaps this isn’t radically implausible when thinking of rogue evil-doers – psychopaths and serial killers and whatnot. But what about Mafiosi? Or perhaps the Nazi family-man, who executes Jews by day and goes home to his loving family at night? The view described here seems to imply that either (a) none of these people have intimate relationships, or (b) they are mutual recognizers. But it seems obvious that (b) is false, since these are all bad people, and we can construct the example such that (a) is false. So it looks like we have an entire class of counterexamples to the position that the good of intimate relationships explains the reason to be moral. While I will, in this chapter, articulate some of the machinery necessary to defuse this worry, a complete defense against the challenge will have to wait until Chapter 6.
any of my own positions to endorse it. Further, then, even if standing to others as mutual recognizer were *necessary* for having intimate relationships, this is not *constitutively* true as was the case with mutual recognition’s constitutive requirement that one act justifiably. Thus, if intimate relationships explain why one has reason to be in relationships of mutual recognition, the nature of this explanation will be different from that in the constitutive case above. It is not the case here that the good of intimacy explains why one has reason to be in relationships of mutual recognition *constitutively*; so in what way does it provide an explanation?

My central thesis attempting to solve these two, related problems (the implausibility of the claim that friendship *requires* mutual recognition and the difference in the *kind* of explanation being utilized) is as follows. We all believe (I claim) that the necessity of performing some act, for the sake of realizing some good, transfers normativity. I want to suggest that the *reliability* of some act in promoting some good also transfers normativity – just less perfectly than necessity. Since reliability comes in degrees, the normative transfer also comes in degrees. Further, then, when normativity is transferred from some good to an act, either by necessity or reliability, the fact of the act’s normative status can be explained by reference to the good. Clearly this thesis needs unpacking, so let us consider some examples.

Suppose that I have a reason to drink water. Suppose further that the only way for me to drink water given the relevant facts of the matter is for me to walk down the hall to the water cooler. It then seems clear that I also have a reason to walk down the hall to the water cooler. The necessity of my so walking ‘transfers’ the normativity from the reason that I have to drink water to the act of walking down the hall. Further, then, this transfer allows me to explain my reason to walk down the hall. If someone demands to know *why* I have a reason to walk down
the hall, I could sensibly say, “Because I have a reason to drink water,” or “Because the good of hydration requires it.” Thus, necessity seems to transfer normativity in a way that enables explanation.

It was this sort of natural reasoning that I appealed to in interpreting Scanlon’s claim of explanation. The fact that an act would be unjustifiable explains why we have reason not to do it, on Scanlon’s view, because acting justifiably is necessary for being in the relation of mutual recognition. This is a genuine explanation because standing in the relation of mutual recognition is ‘worth seeking for its own sake’, which on Scanlon’s theory of value just means that there are reasons to seek, or to be in, such relationships (1998, 96). So we have reason to stand in relations of mutual recognition with others, and acting justifiably is necessary in order to so stand; thus, we have reason to act justifiably.

These observations concerning necessity are intended more as diagnosis than as argument: I take it that this is what we in fact believe about normativity. My claim, then, is that the relation of reliably promoting some good also serves to transfer normativity. When some act X would reliably promote some good Y, then one has reason to X.¹² Further, then, the reason one has to X is explainable in terms of Y. Returning to our water-cooler example, then, suppose that walking down the hall is not necessary in order to hydrate (as I have many other options),

¹² Further, one’s reason to X is stronger or weaker depending on how reliably doing X promotes Y. There are interesting details to mull over here, such as whether ‘completely reliable’ transfers normativity at the same, full force that ‘necessity’ does; similarly, if the relation begins to seem so weak that it seems inappropriate to call it ‘reliably’ promoting, then we may start to wonder at what point ‘very little reason’ turns into ‘no reason’. However, these details need not concern me here. For now, I am simply concerned to point out that the ‘reliably promoting’ relation transfers normativity in a way similar to necessity.
but doing so would reliably realize the state of my drinking (as the cooler has never let me down before). In this second case, it still seems plausible that I have a reason to walk down the hall, and that this reason is explained by the reason I have to drink water. Just as the necessity of performing some action for the sake of realizing a good transfers normativity, so too does the fact that an act reliably promotes some good.

With this thesis in hand, both of the challenges raised at the beginning of this section can be answered. First, recall that one might be skeptical that friendship or other intimate relationships require standing in mutual recognition with others. We can now see that even if being a mutual recognizer merely reliably promotes the good of intimacy, then this too allows us to explain one’s reason to be a mutual recognizer in terms of the good of intimacy. Going back to Scanlon’s kidney-thief example, then, I think that this is precisely what is plausible. That we would find such a friend ‘unnerving’ does not show that mutual recognition is required for friendship; however, it does show that we tend, with good reason, to distance ourselves from those uninterested in standing in mutual recognition. Thus, refusing to stand in relations of mutual recognition reliably tends to block one’s access to intimate relationships, and so becoming a mutual recognizer reliably promotes intimacy in a particular way – namely, it removes a block to intimacy. Whereas we might think of the promotion relation, most generally, as the relation of ‘increasing the probability’ of some outcome, the promotion relation here is plausibly a special case of that more general relation: by removing a block to intimacy, one raise the probability from, say, zero, to non-zero; or from very, very low, to at least slightly higher.\footnote{Many difficult questions arise at this point, such as: How reliably does refusing to be a mutual recognizer block intimacy? Does the exact nature of the reliability affect the strength of this...}
One’s reason to stand in mutual recognition with others, then, is explainable in terms of that relationship’s ability to promote access to intimate relationships.

Secondly, one might worry whether, in moving away from a claim of necessity, I am providing the same kind of explanation as Scanlon. We are now in a position to see exactly how the explanation I have offered is both similar to and different from his. Scanlon explains the reason that one has to act justifiably by reference to the good of standing in mutual recognition with others; I interpreted this explanatory claim as one regarding constitutive necessity. The invoking of mutual recognition is able to explain one’s reason to act justifiably because standing in mutual recognition is valuable — worth seeking for its own sake — and necessity transfers normativity. Since acting justifiably is not only necessary, but constitutively necessary for standing in mutual recognition, there is a particularly tight explanation here.

However, the constitutive relation between mutual recognition and acting justifiably is responsible both for the ‘tightness’ or satisfyingness of the explanation, and for the unhelpfulness I pointed to in section 1.3 above. If anything transfers normativity, it is constitutive necessity, but the fact that acting justifiably is an ‘aspect of’ the valuable relation in question is also what makes Scanlon’s solution appear a bit too close to the trivial end of PD. Thus, it should not be

---

reason? Are there cases in which intimacy is not blocked at all by a disinterest in mutual recognition? Because the current stage of the project is relatively modest, I will not attempt to answer all of these questions here. Despite these complications, my hope is that I will have, in this chapter, shown that there is a connection between the general ability to engage in intimate relationships and a commitment to justifiability. Later in the dissertation, however, I will use this argument to explain the existence of the reason to be moral, and the questions listed above will come back with a vengeance (see esp. Chapter 5). At that point, there will be no dodging them, and so I will postpone the long, difficult answers to them until then. Thanks to Steve Kuhn, in particular, for helpful discussion on these points.
surprising that an explanation sought for its increased helpfulness will lack the tight connection between the *explanans* and the *explanandum*.

1.8 Erring According to Relevance

If the arguments above are on track, then whether they would be embraced by Scanlon as being helpful or not, I have shown that some additional, interesting conclusions can be derived from his framework. According to Scanlon, one’s reason to act justifiably can be explained by the value of standing in relations of mutual recognition. However, since standing with others in mutual recognition *just is* to live with others on terms they could accept, the explanation picked out is not sufficiently helpful. Anyone asking the question, ‘What reason do I have to be moral?’ is likely to be unsatisfied with an answer that utilizes the very concepts that form our account of morality.

Fortunately, Scanlon’s own arguments regarding the compatibility of morality with other goods provide a series of claims that I think provide a second candidate view. Friendship, Scanlon claims, requires that one live in mutual recognition with others, because a blatant disregard for the standards of justifiability unnerves those with whom we would like to share a friendship. Although I admitted that this claim of necessity is implausibly strong, I suggested that Scanlon’s basic insight here could still provide the basis for an explanation of one’s reason to be a mutual recognizer, so long as we expand our idea of what constitutes normative explanation. So, just as some act’s being necessary for the realizing of a good provides an explanation of why one has reason to perform that act, so too does some act’s reliably promoting some good. Thus, the fact that standing in mutual recognition with others reliably promotes intimate relationships allows the good of intimacy to explain why one has a reason to stand in
mutual recognition with others. So one has reason to act justifiably because doing so is an aspect of living with others on terms they could accept, which reliably tends to promote the having of intimate relationships with others. Relying primarily on Scanlon’s own insights, then, we have uncovered a chain of reasoning which allows us to explain one’s reason to act morally in terms of the good of intimacy.

We thus have two options for explaining the motivational basis of a contractualist account of morality. I claimed above that Scanlon’s preferred explanation is weak because it errs on the unhelpful, or trivial, horn of PD. Further, we can understand the nature of this unhelpfulness by looking at Scanlon’s two components of moral motivation: a successful theory must explain both what those who are gripped by morality are moved by when they act morally, and why everyone has a reason to care about being so moved. My charge is that, while Scanlon’s account discharges the first burden, it fails to satisfactorily discharge the second. By explaining the reason to act morally by reference to a moral relationship, we haven’t cast the motivational net any wider than it was already cast: those who independently care about morality will be moved by such considerations, while those who were in a position to genuinely ask the question are likely to be left cold. In order to move away from the unhelpful horn of the dilemma, we must explain the reason to be moral by reference to a good that is both sufficiently detached from morality to engage those asking the question, and by reference to something so recognizably good that it becomes plausible to think the net has really been cast more widely.

The good of intimacy satisfies both of these requirements: it is a non-moral good, and so does not presuppose that one is already interested in morality, and it is obviously and recognizably valuable. Thus, by explaining the reason to be moral ultimately in terms of the
good of intimacy, I have clearly discharged Scanlon’s second burden, as it is clear in this case why one ought to care about the reason to act morally – because that reason is grounded in intimacy, which is clearly valuable.

The worry at this point, however, is obvious. PD predicts that, having satisfactorily explained why everyone has reason to care about acting morally, I must have erred on the side of offering implausibly external incentives for being moral. And thinking about the account, this seems quite possible. On my suggested account, one has reason to be moral because doing so promotes intimate relationships. And this may, in fact, seem a bit too much like saying that one has reason to be moral because doing so promotes one’s own happiness. Surely, the critic will say, acting rightly because doing so promotes friendship is the wrong kind of reason for moral action.

I might defend the intimacy-based account by saying, “sure, my view is on the ‘less morally relevant’ end of the spectrum, but I haven’t erred to the same extent as the most egregious offenders. After all, I have explained the reason to be moral by reference to valuable relationships, and surely this is more morally relevant than one’s happiness.” And in fact, I think this response is accurate. However, we have seen that Scanlon also doesn’t err as egregiously as the most trivial views, as Scanlon’s explanation included reference to a relationship, and so wasn’t simply explaining the reason to be moral by reference to morality. So if it is true that my proffered solution has done no more than move slightly towards the center from one horn of PD, then this does not militate in favor of my solution over Scanlon’s, as his has done that as well.

1.9 (Really) Solving Prichard’s Dilemma
At this point, I have shown both: (1) that the space between the horns of PD really is a spectrum, as it makes sense to say that positions succeed or fail to account for the competing desiderata more or less; and (2) that Scanlon’s contractualism leaves us with at least two options for explaining the motivational basis of morality. And this is interesting by itself, as it gives one options for choosing which side of the spectrum seems more plausible. However, I do not think I must conclude quite this modestly. What I want to suggest here at the end is that the above investigation serves to support the difficulty of PD in a way, as it shows that moral motivational accounts specifically designed to solve PD still struggle with fully discharging its burdens. This, I think, lends support for the strategy that I want to pursue, which is to require that the moral motivational story be complex, meaning that it provides multiple explanations of one’s reason to act justifiably. While some of the arguments gestured at here will write checks to be cashed later in the dissertation, I hope that even in skeleton form, their persuasiveness will be evident.

The first thing I want to do is offer the candidate moral motivational views thus far surveyed as evidence of the genuine difficulty of perfectly solving PD. The challenge is introduced by reference to actual views that make up the most radically opposing answers. On the one hand, most of us (I think) at least sometimes feel the pull of the Kantian intuition that one has reason to act rightly simply because doing so is her duty. Although we don’t always respond to this reason when doing the right thing, it seems that paradigmatically moral actors are often moved by consideration of their duty. And on the other hand, Socrates’ interlocutors in the Republic are easy to sympathize with when they demand to know how it is always in one’s interest to be just, as a moral motivational story based on one’s interest or happiness seems like perhaps it would be the most satisfying kind of answer.
In trying to find a solution that was less radical than either of these initial positions, we have succeeded: Scanlon holds that the reason to be moral is explainable in terms of the good of mutual recognition, and I suggested that Scanlon’s own arguments could also be marshaled to suggest that the reason to be moral is explainable in terms of the good of intimate relationships. Now each of these views fares better than the radical positions that I started with, but both still, I suggested, seem unsatisfactory, for the same reasons as the radical positions – just to lesser degrees. So Scanlon’s solution is ‘less trivial’ than an explanation based on duty, but the worry is that it still fails to be genuinely helpfully explanatory. And my offered solution is ‘less morally irrelevant’ than an explanation based on one’s happiness, but may still seem to provide the wrong kind of reason to act rightly.

We should be unsurprised at this result. Scanlon said in an earlier quote that “[a]nswers [to PD] can thus be arrayed along one dimension according to their evident moral content, ranging from those that appeal to what seem most obviously to be moral considerations…to those having the least connection with moral notions” (1998, 150). When demanding from a view that it be both morally relevant and not too tightly connected with morality, then, we should be unsurprised if answers that employ moral concepts seem to fail according to one standard (to some degree) while those that employ only non-moral concepts seem to fail according to the other (to some degree). It thus looks like we may be stuck arguing the merits of falling on one side of the spectrum vs. the other, as all explanations will either employ moral concepts or employ none.

There is another option, however, and it is the one that I want to sketch here. Since PD challenges any account of moral motivation to satisfy competing desiderata, one could give up
trying to satisfy both at once, and instead offer a complex solution with two component parts – one for each desideratum. Now certainly, if the only way to accomplish this were completely ad hoc, that would be a (likely decisive) strike against the view. But that is not the case. In fact, the view that I offered based on Scanlon’s arguments – the intimacy-based account of moral motivation – is precisely such a view, when understood correctly.

The intimacy-based view is not most promising because its errors are less egregious than Scanlon’s, or because it is more important to point to a clearly valuable good in explaining moral motivation than to employ a morally-relevant consideration. Rather, the main reason for preferring the intimacy-based view is that it includes Scanlon’s view. Recall that I did not merely claim that one has reason to act rightly because one has reason to be in intimate relationships, although I did summarize the view this way. There is a middle term in the explanatory story, and it is Scanlon’s own term. One has reason to act rightly because one has reason to be a mutual recognizer, and one has reason to be a mutual recognizer because one has reason to be in intimate relationships. Thus, the intimacy-based view entails that the complete explanation of the motivational basis of morality is complex, and involves both a clearly moral component and a component with obvious, recognizable value. If Scanlon is right that a full moral motivational story must explain both why those who are gripped by morality are so moved and why everyone else has a reason to care about being so gripped, then we are faced with a challenge that can be fully answered only with a complex explanation, and here we have one.

Now one might respond that the middle term in the explanatory story is excised in the summary for good reason, which is that it does not get to do any heavy lifting. The ultimate reason for acting rightly is what matters, and so my view does not get the benefit of utilizing
both reasons. Instead, an explanation in terms of the good of intimacy is impaled on the ‘external incentives’ horn, perhaps only slightly less severely than explanation in terms of one’s happiness.

I do not, however, think this is right. In fact, part of what I find so inspired in Scanlon’s move to relationships is precisely its ability to solve this problem. Whatever we refer to as the ultimate reason to be moral needs to be a powerful explanation of why everyone has a reason to care about being moral, but it also, I think, needs to explain not why one should act rightly, but rather why one should be a certain way. While I will have a chance to explain the importance of this move in a broader context in Chapter 5, let me here explain its importance for moral motivation.

It does seem that we want an agent’s reason for acting morally to be morally relevant. To see this, we need only conjure up a toy case: if a man saves a drowning infant, and we ask him why he did so, we would be deflated to hear him respond, “Well, my happiness was at stake; it would have really bummed me out for a few days to have seen a baby die.” And while I have contended that an explanation by reference to relationships fares better, a mirror case might generate similar intuitions. If we ask a woman why she refrained from murdering in cold blood, we might be disheartened to hear her respond, “Well, because murdering someone is likely to affect my relationships with my partner and friends.” But our intuitions here do not tell us that we cannot explain the reason to be moral ultimately by reference to something that stands in a rather distant relationship to morality; rather, they tell us that we don’t want those distant reasons to be what are consciously moving us when we are acting.
If the above diagnosis is correct, then PD only seemed like a dilemma because of an ambiguity in the language of ‘the reason’ to be moral. Scanlon himself, we noted, saw that a moral motivational story had to be complex, featuring both motivating and justifying components, but he did not see that the answer might yield different parts of the theory to shoulder those respective burdens. That one ought to be motivated (at least sometimes) by morally relevant considerations does not entail that those same considerations must be the ultimate explanation of the reason to be moral. And just because we want an account that makes sense of why acting rightly should seem valuable to everyone (which suggests a wide-reaching, non-moral value) does not mean that such a consideration ought to be what actually moves one to action. Scanlon’s keen diagnosis of the challenge facing accounts of moral motivation wears on its face a suggestion for how to meet that challenge: there are two desiderata, and they are best met by employing two different aspects of a story, and so the motivational basis of morality should be complex.

1.10 Conclusion

In this chapter, I have tried to motivate Scanlon’s challenge to theories of moral motivation – what he calls Prichard’s Dilemma – and to investigate various ways to solve it. Both the solution that Scanlon offers, and a solution that I offer on his behalf, fail for similar reasons. While Prichard’s Dilemma is not a genuine dilemma, it does challenge any theory of moral motivation to satisfy competing desiderata, and so candidate solutions do seem to err on one side or the other, even if they do not all do so to the same extent. However, I also closed by sketching what I take to be the form of a genuine solution to the ‘dilemma’, which requires
coming up with a story sufficiently complex so as to satisfy the differing criteria of success with
different component parts of the theory.

I then offered my own complex view, which holds that one has reason to act justifiably
because one has reason to be a mutual recognizer – a reason which one has because one has
reason to be in intimate relationships. The good of intimacy thus explains the reason to be a
mutual recognizer, which explains the reason to act justifiably; and it may be the case, in
addition, that paradigmatically moral actors will often act rightly only on the basis of
considerations of mutual recognition – that is, moral actors may take as their reason for acting
that doing so is an aspect of a positive way of living with others. There is nothing incompatible
in having, as one’s ultimate reason for acting rightly, a value that is only slightly relevant to
morality, while maintaining our intuitions that paradigmatically moral actors ought to act on the
basis of morally relevant considerations.

With this account of moral motivation in hand, I will now turn to a more revisionary
project: namely, providing an existential (as opposed to a merely motivational) basis of morality.
In addition to helping to solve Prichard’s Dilemma, identifying intimacy as a non-moral good
capable of generating a reason to act justifiably also points the way towards a novel metaethical
basis for contractualism. Since intimate relationships are the kinds of thing people generally
desire and care about, all we need is a theory of normativity that explains the existence of
reasons in terms of what people care about. But of course, there is such a view, which goes by
the name of Humeanism. The work in this chapter, then, will provide the foundation for
developing a Humean grounding for Scanlon’s Kantian contractualism.
CHAPTER 2

In the previous chapter, my goal was to critique and correct Scanlon’s account of the motivational basis of morality. More colloquially, I asked and attempted to answer the question ‘Why be moral?’ (call this the ‘WM’ question) on behalf of Scanlon’s view. However, I asked the question in a restricted sense. On the one hand, the WM question seems to ask a ‘which’ question: which reason, among some set of reasons, is the one that does the moral motivational work? This ‘which’ reading of the question is restricted because it assumes that there is, in fact, a reason to be moral. However, it is also easy to hear the WM question in a broader way. On what I will call the ‘existential’ reading, the WM question asks whether there exists a reason to be moral. The most complete answer to the WM question will respond both to the restricted and the existential readings. That is, a complete account of the reason to be moral will explain both the existence of that reason and its link to motivation.

Scanlon did, I think, provide a complete account of the reason to be moral in What We Owe to Each Other, as he dedicated much of the beginning of the book to defending his ‘primitivist’ view of reasons. As we will see in more detail below, primitivism about reasons is a form of metanormative ‘quietist’ realism, according to which some thing’s being a reason cannot be helpfully or non-circularly explained (1998, Ch. 1). On this view, all that can be said to explain the concept of a reason is that it ‘counts in favor of something’; but to count in favor of something, then, is simply explained as the ‘being a reason for’ relation (1998, 41).

Note that primitivism holds only that what it is to be a reason cannot be helpfully or non-circularly explained; that some particular fact, R, is a reason for me to phi, often will admit of some explanation. Scanlon’s concern, then, is not with explaining particular reasons, but explaining the concept or property of being a reason.
Although he does not say so explicitly, it seems plausible that on Scanlon’s view, reasons to be in intimate relationships and relations of mutual recognition exist, as such relationships are clearly valuable – there are all sorts of features of such relationships, such as the satisfaction that they tend to promote, that count in favor of each of us pursuing such relationships. So long as this is true, then Scanlon has given us an account of reasons on which we can understand why there exist the particular reasons discussed in the previous chapter, and the arguments there show how such reasons relate to moral motivation.

Although Scanlon has, in fact, provided all of the pieces to generate a complete account of the reason to be moral, I do not think the resulting view is a particularly satisfying one. In fact, I will argue in 2.2 that the ‘unsatisfying’ nature of his account of the reason to be moral is precisely what is responsible for one of the most widespread (and often uncareful) criticisms of his view. I will then go on, in 2.3 and 2.4, to show how one might take a cue from Scanlon’s earliest articulation of his contractualist theory in order to provide a more satisfying account. In short, I will argue that Scanlon’s ‘constructivist’ moral theory is ill-served by a primitivist metanormative theory, and that his view would be strengthened if he were to revert to something closer to what I call his ‘Humean Kantianism’ of the 1980s. Finally, in section 2.5, I will formulate the basic claims that constitute my own Humean Contractualist view, and make some initial observations about the strengths and weaknesses of such an account.

2.1 Primitivism about Reasons

According to Scanlon, the idea of a reason does not admit of any non-circular explanation. To be a reason is to be a consideration that counts in favor of something. The ‘counting in favor of’ relation, however, is the kind of thing one might want further explained,
and Scanlon thinks that all we can say is that a consideration counts in favor of something by being a reason for it.² The idea of a reason, then, is primitive on Scanlon’s view (1998, 17).³

Although Scanlon does not introduce the view this way, primitivism seems to be a non-naturalist, realist view of reasons. Claims concerning what facts constitute reasons to do which things are made true by those facts counting in favor of some agent’s doing something or other. Parts of the world – propositions or facts, on Scanlon’s view – just do sit in the ‘counts in favor of’ relation to other parts of the world (1998, 17). But as Sharon Street notes, primitivism, thus understood, is best read as a non-naturalist, realist position (2008, 218 f. 25). On Scanlon’s view, reasons are fundamental, unanalyzable, irreducibly normative concepts.⁴

² Further, Scanlon claims that his reluctance to explain what it is to be a reason both does not leave undone anything that needs done, and is likely compatible with anyone who wants to provide a substantive explanation of what it is to be a reason. That is, Scanlon holds that the idea of a reason is not a problematic one in need of any explanation, but also that nothing he argues depends on the failure of other, more explanatory accounts (1998, 17). As I will argue in this chapter, I think he may be wrong on the first count – it may be that his view, in particular, needs a better explanation of what it is to be a reason; however, the goal of this dissertation is to vindicate the second claim – that his moral theory can successfully be married to a different, more explanatory account of reasons.

³ Scanlon later (2014) distinguishes multiple ways in which one can take reasons to be ‘primitive’ or, as he comes to say, ‘fundamental’. One such way is simply to hold that there is no non-circular explanation of what it is to be a reason; reasons cannot be reduced or further analyzed. Another way to hold that reasons are primitive or fundamental is to claim that they are the basic normative unit, and so constitute all other normative facts. Although Scanlon admits that he believes this latter form of reasons primitivism, he explicitly defends only the former view (2014, 2).

⁴ Although Scanlon has now come to accept the label of ‘nonnatural realism’, he is careful to reject the idea that realism implies an ontological commitment. In his most recent thinking on the topic (2014), Scanlon leans on the distinction between analyzing a concept (what he is doing) vs. analyzing a property (which he is not doing) in order to avoid charges that a realist position like his is ‘queer’ in any way. I will have more to say about this maneuver later. For now, simply note that the ‘realism’ moniker tends to be used (and has been used in reference to Scanlon) to identify views that require an expanded ontology – one, for instance, that implies the existence of a normative property of being a reason. For this reason, Scanlon’s revised view (in
Like Moore’s account of the good, Scanlon’s view of reasons is not simply non-naturalist and realist, however; it is also, to adopt the language of Ralph Wedgwood, ‘quietist’ (2007). According to Wedgwood, what makes a view quietist is its commitment to realism, but also to the claim that there is no substantive or illuminating explanation of normative reality. Because the quietist cannot say anything substantive or illuminating about normative reality, he is left diagnosing the mistakes of opposing views.

There are many, well-worn criticisms of ‘realist’ views. J. L. Mackie famously argues that non-naturalist views like Moore’s are metaphysically ‘queer’, as they do not fit anywhere into a naturalistically plausible ontology (1977). Relatedly, because of the queerness of irreducibly normative entities, one might worry that such realist views will require objectionably ‘intuitionist’ epistemologies, or that such facts will be explanatorily otiose. Such criticisms have been influential for decades, and I in fact find all of them compelling. However, they are a

---

5 Paradigm cases of quietism include, according to Wedgwood, McDowell (1998) and Parfit (2011).
6 For the most straightforward case of this methodology, see (Parfit 2006). Scanlon, too, seems to embrace this strategy in his full defense of realism (2014), organizing the chapters of the book around an argument from elimination. At the end of the introductory chapter, Scanlon claims that, “If [explanations of reasons in terms of desire and rationality] are unsatisfactory, and if, as I will argue in the next two lectures, the appeal of expressivist views should be resisted, then unless there is some other general account of reasons (which I doubt) Reasons Fundamentalism [Primitivism] will be left as the only available option” (2014, 14).
7 See (Wedgwood 2007) for an argument that realism requires intuitionism (but also for an argument that intuitionism is not, in fact, all that objectionable).
8 For the now-classic argument, see (Harman 1975); see also Nicholas Sturgeon’s powerful response in his (1985).
bit out of fashion, with long debates in the literature proving that they will not convince everyone. Further, there seems to be a bit of resurgence in normative realism as of late, with many new realist positions being defended in the literature, none of which seem bothered by these classic worries and all of which are bolstered by interesting, thought-provoking arguments.\(^9\)

I do not know how to adjudicate among the arguments in this debate. My own intuitions are that the classic worries about queerness, intuitionism and otiose explanation are in fact sufficient to reject any view that posits non-natural, normative properties,\(^10\) but such a conclusion is clearly not the only reasonable one to draw from the dialectic. In addition, such worries are precisely what drove Scanlon recently to attempt a ‘non-ontological’ defense of his realist position (as mentioned above in f. 4). On his most current thinking, a view like his can be ‘realist’ without entailing any claims about normative properties (2014, esp. Lecture 2). Thus, even if queerness-style objections are, in fact, decisive, if Scanlon’s defense works, at least one version of primitivism will be immune to such concerns.

For these reasons, I will not here criticize Scanlon’s primitivist view on ontological grounds. Instead, I will criticize his view on explanatory grounds. In particular, I will argue that one of the most wide-spread criticisms of Scanlon’s contractualism is in fact due to the

---


\(^10\) I take it this is unsurprising, given my commitment to philosophical modesty (discussed in 0.1). However, as I will go on to discuss here, I think that there are good reasons to reject Scanlon’s primitivism without consideration of ontological modesty, and so I will attempt to mount my case for rejecting this aspect of his view without relying on these metaphysical and epistemological arguments.
unsatisfying nature of his theory of reasons *given his account of morality*, as the introduction of a theory of reasons to undergird contractualism promises an explanatory pay-out that it never realizes. My conclusion will then be that, regardless of whether primitivism does or should make any ontological claims, there is especially good reason to reject such a view when paired with Scanlon’s contractualism.

2.2 The Circularity Worry (and What it Reveals)

One of the most common criticisms of Scanlon’s contractualism to emerge in the wake of *What We Owe to Each Other* is that the view is ultimately circular. Although foreseen by Scanlon, who attempted to defuse the worry ahead of time (1998, 195-223), this criticism received quite a lot of play, both in book reviews\(^\text{11}\) and scholarly articles.\(^\text{12}\) Now, my own view is that it is quite clear that Scanlon’s view is *not* circular, and that the most common form of this charge results from a sloppy reading of the position. However, I also think that the circularity worry points to an understandable unease with the position. In short, I will argue that the widespread dissatisfaction with Scanlon’s view, often expressed in the form of a ‘circularity worry’, in fact points to a concern with Scanlon’s constructivism/realism pairing. Making this case is the goal of the current section.

Even on the surface of Scanlon’s contractualism, it is easy to see how a circularity worry gets started. First, consider Scanlon’s criterion of wrongness: an act is wrong if its performance would be disallowed by any set of principles that no one could reasonably reject. Now within this criterion is the standard of reasonable rejection, and so we will want to know what counts as

\(^{11}\) See Blackburn (1999) and McGinn (1999).

\(^{12}\) For just a few examples, see Hooker (2003), Pettit (1999) and Ridge (2001).
reason. If I reject a principle that disallows theft because I like free stuff, presumably my rejection would not be counted as reasonable, and so would not count against the standing of the candidate principle. But why is my rejection unreasonable? Well, here is the challenge. If we count my rejection as unreasonable because it seems morally problematic – which does seem to be the problem – then the contractualist machinery actually presupposes moral standards. Or, as Colin McGinn puts the point:

The first and most pressing problem with Scanlon’s theory is that it is circular. It presupposes what it is meant to explain, in the most obvious way. An action is said to be wrong if it may by criticized by others. But on what grounds is it criticizable? On moral grounds, surely. But then aren’t we back where we began? An action is wrong if it can correctly be said to be wrong. But weren’t we being told what it means to say that an action is wrong? Circle! (1999, 35)

More carefully, we can follow Michael Ridge’s dilemmatic formulation of the challenge (2001): a reason to reject some principle is either a moral reason or it is not. Now if it is not a moral reason, it seems unfit to the task. That some principle would cause me unnecessary pain, by itself, seems like a bad reason for rejecting shared principles to live by. It would be a good reason only if my pain were morally relevant. Or, to use Blackburn’s example: that some principle would allow vast inequalities of wealth is a good reason to reject some principle only if vast inequalities in wealth is morally bad (1999). So, the argument claims, the reasons to reject

13 Although Ridge provides a particularly clear summary of the worry, I should note that he does not find the concern moving. His goal in the cited article is to ‘save Scanlon’ from the circularity worry.
principles must be *moral reasons*. It may well be the case that there are good reasons to reject principles which allow vast inequalities in wealth or the causing of unnecessary pain, but those reasons are based in the moral badness of those states of affairs. But if this is the case, then contractualism is in trouble. Its job was to explain moral reasons, but it is capable of doing so only by appealing to moral reasons. If we must suppose such reasons anyway, then it appears that the contractualist machinery is a spare wheel, spinning but bearing no weight in the moral theory.\(^{14}\)

Although I will shortly argue that such concerns do, in fact, point to a real worry for Scanlon’s view, there is no legitimate circularity challenge. Scanlon’s contractualism has the job of explaining the existence of moral reasons. Let’s stipulate that a moral reason is a reason given to an agent in virtue of some act’s status as obligatory or impermissible. So the fact that an act would be disallowed by any set of principles which no one could reasonably reject gives all agents a moral reason not to perform that act. Scanlon’s contractualism would be circular, then, if it helped itself to moral reasons.

However, it does not. Scanlon’s contractualism requires that there exist *reasons* – but it does not presuppose moral reasons. In fact, the correct way to understand Scanlon’s contractualism is that it is precisely what is required in order to turn general reasons – fully normative, but non-moral considerations – into moral reasons.\(^{15}\) Since morality, for Scanlon,

\(^{14}\) The spare wheel analogy is Hooker’s (2003).
\(^{15}\) I take it that this is similar to the point that Sharon Street makes about ‘restricted constructivism’ – a class of views that includes Scanlon’s – when she states that there is no reason in principle why a version of restricted constructivism must be circular. In particular, the mere fact that a restricted constructivist view characterizes some reasons in terms of others does not imply that it is circular, for the *goal* of restricted views – and
centrally concerns living with one another on terms that all can accept, general reasons are turned into moral reasons by being fed into the contractualist criterion and seeing which considerations survive. All of us have reasons to do things, to prefer things, to want to live in certain ways, but only some of them will facilitate our living together on mutually acceptable terms. The very job of the contractualist machinery is to take the entire set of general reasons as input, turn the crank (consider which ones are acceptable to all others), and generate as output the set of moral reasons.

Now of course, there is a challenge here: Scanlon must be able to articulate which reasons for principle rejection are legitimate – that is, what is a reasonable rejection of some principle. And doing this in a way that doesn’t import any moral considerations is tricky. However, Scanlon has an answer to this challenge, resting on his delineation of ‘generic reasons’, or “reasons that we can see that people have in virtue of their situation, characterized in general terms, and such things as their aims and capabilities and the conditions in which they are placed” (1998, 204). Although the job of making clear the content of such generic reasons is not easy, there is no reason to assume that Scanlon has begged any questions here. As a primitivist, he is entitled to presume the existence of a set of reasons, and he does. These reasons, then, are the basic considerations that are brought to the contractualist table; each person must leave behind the idiosyncratic ones, and then the job of the contractualist machinery is to determine

this point is critical to understanding them – is not to provide a characterization of reasons in general, but rather to provide a characterization of some limited subset of reasons, such as reasons of political justice or reasons of right and wrong...to avoid circularity, a version of restricted constructivism must simply avoid including the results of construction among its materials of construction; it must ensure that there is no overlap between its target set and its grounding set of normative judgments (2008, 215).
what falls out of this process. And this sounds exactly like how a contractualist procedure should go, as it takes considerations of individuals and attempts to generate interpersonal obligations out of them.

Despite my defense of Scanlon on this point, I do think that there is a legitimate concern with Scanlon’s view somewhere in the neighborhood. In addition to worrying about a ‘circle’ or that Scanlon begs the question, some commenters have formulated the worry as the charge that Scanlon’s view is ‘redundant’, as the contractualist machinery is an unnecessary addition to ordinary moral reasoning.\(^{16}\) What I want to suggest is that all such worries are generated because Scanlon’s contractualism presupposes his primitivist view of reasons.

When a critic is trying to make sense of the contractualist criterion of permissibility, it is noted that there is normative language in the *analysans* – namely, part of what makes an act impermissible is that the principles that allow it can *reasonably* be rejected. And because Scanlon is a primitivist, that normativity is never eliminated. The reasons that are given in objection to some principle sit at the bottom of the analysis. As I argued above, Blackburn, McGinn and others are wrong to think that these are *moral* reasons, and so the view isn’t circular. However, there is something unsatisfying about the conversation stopping here. The criterion of permissibility tells us something about morality, and then right when it gets interesting – right when we want to know what is wrong with some set of principles – it looks as

\[^{16}\text{See Pettit (1999), Stratton-lake (2003).}\]
though Scanlon is left simply assuring us that there are reasons to reject them. It seems natural to want to know why there are such reasons.\(^{17}\)

My diagnosis of this dissatisfaction can be expressed in two, compatible ways. First, note that contractualism is a *constructivist* view. According to what might be taken to be the canonical view of constructivism, a view is constructivist if it holds that a hypothetical procedure determines the principles of morality (Darwall, Gibbard and Railton 1992); however, Sharon Street argues convincingly that constructivism need not be hypotheticalist, but rather, that it must hold that normative truth is determined by what is entailed from *within the practical point of view* – that is, from within the view of already-valuing agents. For my purposes, then, I will label a view constructivist if it shares the commitment that there are no moral truths or facts independent of either (a) some hypothetical procedure, or (b) the practical standpoint of agents (Street 2009).

Constructivism entered the moral lexicon with John Rawls, first in the political context (1971) and then in the moral context (1980). And while ethical constructivism (like Scanlon’s) is more ambitious than narrowly political constructivism (like Rawls’ in (1971)), the two positions share the view that normative reasoning begins from within the practical point of view,

\(^{17}\) I think that Street’s reaction to constructivist views in general seems to support my hypothesis. As noted in f. 9 above, Street saw the circularity worry as obvious enough to insert the claim that constructivist views like Scanlon’s do not *need* to be circular, although they might be if they aren’t careful. As will become clear below, I take it that Street and I are playing for the same team here, as we are both interested in rejecting ‘restricted constructivist’ views like Scanlon’s in favor of a view with more explanatory power. In her case, as we will see, this involves generating a ‘thoroughgoing’ constructivist position; whether my own view ends up fitting into this class is unclear, but regardless, our motivations for rejecting restricted versions like Scanlon’s seem to be the same.
and so is a *practical* problem. For Rawls’ position, this meant the ability to move forward with normative reasoning *without* being forced to solve every sticky metaethical problem raised in the first half of the 20th century. Normative constructivism, then, offers a kind of permission to do good, first-order normative reasoning without getting mired down in the metanormative issues.  

Although constructivism, so understood, offers a kind of freedom, it also leads to a kind of unease for those philosophers who are interested in finding a complete explanation of morality. So it is understandable that a moral or political philosopher may come to a constructivist position and want to know *what explains* the normativity of the considerations being fed into the constructivist procedure. This question may have been what led Scanlon to abandon a narrowly constructivist project and to undergird the first-order moral theory with a substantive theory of reasons. However, it is also, I suggest, what led Scanlon’s critics to look for an explanation of the reasons that were being fed into the contractualist machinery. The problem, then, is that Scanlon makes a move that promises an answer – he seemingly admits that he needs an explanation of the normativity being fed into the constructivist procedure – but he then provides us with an answer that feels like a non-answer. Appealing to a realist theory of reasons, about which we cannot say anything helpful, moves outside of the constructivist framework without providing any real explanatory benefit.

Or, on the second way of understanding the dissatisfaction: if Scanlon is willing to import a realist theory of reasons, why not get it to do all of the hard work? Primitivism about reasons

---

18 Although he hadn’t begun using the language of constructivism yet, see on this point Rawls’ (1951). On my reading of the history, Rawls’ constructivism emerged from a frustration with the seeming stalemate in metaethics, brought on by Moore’s (1903) and the challenges of the non-cognitivists like Ayer (1936) and Stevenson (1937).
already requires that Scanlon face all of the major criticisms of realism – the metaphysical extravagance, the epistemological worries and the like. The question, then, is why anyone should take on such baggage and then attempt to motivate a contractualist criterion of wrongness. While it is wrong to say that Scanlon imports moral reasons, one might reformulate the circularity and redundancy objections as a demand to know why we shouldn’t import moral reasons. So long as we’ve accepted that reasons exist, and that there is no informative, non-circular explanation of such reasons, it makes the most sense to take what look like moral reasons (like the fact that some principle would allow vast inequalities of wealth) to be, in fact, moral reasons, and so to dispense with contractualism.

This concludes my diagnosis of the circularity and redundancy worries for Scanlon’s contractualism. While they do not, in any of the forms I have seen, succeed in showing that the view fails, such worries do point to an understandable dissatisfaction: namely, that if a view promises an explanatory pay-off, it shouldn’t stop the conversation too soon; and if a view takes

19 One might think that the answer here is rather straightforward, and it is that there is a difference between reasons and moral reasons – namely, that the former is more basic than the latter. While I think this is the best answer Scanlon can give (and is the one I give on his behalf above – that reasons are what go into the ‘contractualist machinery’ to generate moral reasons), the criticism here is orthogonal to the question of whether there is a relationship of ‘more basic’ and ‘less basic’ among various reasons. The worry here is that one is paying (what many take) to be a theoretical cost by being a constructivist, by explaining moral reasons in a proceduralist or practical way. But if one is willing to take on the baggage of non-natural realism, the challenge is that she will be provided a more satisfying explanation of moral reasons by appealing to her realism than by appealing to constructivism. Yes, perhaps moral reasons will still turn out to be different from other reasons, but constructivism constrains one in how she explains this difference. The challenge here, then, is that one who already has realist commitments will do better to explain the difference by reference to those commitments than by building them out of constructivist materials. As a sound-bite: a goal of much constructivist reasoning is to avoid realism; once you’ve embraced it, why bother with constructivism? Thanks to both Steve Kuhn and Mark Murphy for helpful discussion on this point.
on the baggage of realism, it ought at least to vindicate our realist intuitions. Given that these are conditional imperatives, they could be satisfied by a contractualist view that either (a) stayed out of metaethics altogether; or (b) utilizes a more appropriate metanormative view. As I will explain below, I believe that Scanlon should have taken the latter route. Even in its earliest formulation, Scanlon’s contractualism was already reaching for a substantive metanormative view that would give it some explanatory power. In the remainder of this chapter, I will argue that this instinct was in fact a good one, and that rigorously following through on this early attempt results in a moral theory that is more satisfying as a complete package.

2.3 Scanlon’s Humean Kantianism

The first published articulation of Scanlon’s contractualist view appeared in the article “Contractualism and Utilitarianism” (1982), more than 15 years prior to the publication of *What We Owe to Each Other*. It should be unsurprising, then, that several features of the view changed in the intervening years, as Scanlon honed his position. What I want to point out in the current section, though, is that Scanlon himself first articulated his view in a non-primitivist way, and that this earlier version of his view is unlikely to have led anyone to have the circularity or

---

20 As I will discuss later, views that are more ‘appropriate’ might come in two different flavors. On the one hand, one might be open to trying a thoroughly naturalist, reductionist project, as such a view promises broad explanatory power. On the other hand, one might, with Sharon Street, hold that what is needed is a ‘thoroughgoing constructivism’, or a view that applies the constructivist procedure both to the first-order moral theory and the grounding metanormative view. Whether these views will always differ is unclear, as ‘Humeanism’ about reasons (the view I will defend later in this chapter) is a reductionist (realist!) project in the hands of Mark Schroeder (2007), but ‘Humean Constructivism’ is a similar, but procedure-based account of reasons (Street, 2008). For further discussion of the difference between constructivism at the level of first-order moral theory and at the metanormative level, see (Korsgaard 2003) and (Schafer 2012).
redundancy objections above. While I will not argue that Scanlon’s early view was exactly correct, I do want to point out the virtues of the form that the view took. Although Scanlon eventually abandoned it due to perceived difficulties, in the remainder of the chapter I will argue that he could have tweaked the view rather than abandoning it. If he had, he would have improved on the problematic elements of the first iteration, without inviting the worries that came along with primitivism.

As I showed in Chapter 1, Scanlon goes to great lengths to provide an account of the motivational basis of morality, and that same concern is visible in his first articulation of the view. However, in “Contractualism and Utilitarianism” (henceforth: CU), Scanlon did not yet show the nuance of analysis that I canvassed in the previous chapter. He did not explicitly recognize the competing justificatory and motivational elements of the moral motivational question, and he did not have a primitivist account of reasons on which to rely. As a result, his account of the motivational basis of morality was quite simple: since wrong acts are ones that cannot be justified to others on grounds that they could accept, moral motivation is provided by the widespread desire to act justifiably (1982, 276-277). And, because he was not yet as careful with the account as he ultimately would become, this explanation was intended to be both motivational and justificatory.

Evidence that this desire was intended, by Scanlon, to do both motivational and justificatory work is provided by claims made within just a few pages of one another. First consider the claim that “the source of motivation that is directly triggered by the belief that an action is wrong is the desire to be able to justify one’s actions to others on grounds they could
not reasonably reject” (276). Then notice that only a few paragraphs later, Scanlon tells us that one of the virtues of contractualism is that it rests in part on a qualified skepticism. A non-contractualist theory of morality can make use of the source of motivation to which contractualism appeals. But a moral argument will trigger this source of motivation only in virtue of being a good justification for acting in a certain way, a justification which others would be unreasonable not to accept. So a non-contractualist theory must claim that there are moral properties which have justificatory force quite independent of their recognition in any ideal agreement. These would represent what John Mackie has called instances of intrinsic ‘to-be-doneness’ and ‘not-to-be-doneness’. Part of contractualism’s appeal rests on the view that, as Mackie puts it, it is puzzling how there could be such properties ‘in the world’. By contrast, *contractualism seeks to explain the justificatory status of moral properties, as well as their motivational force, in terms of the notion of reasonable agreement* (1982, 277; italics are mine).

What this long quote reveals is that Scanlon saw contractualism’s rejection of Moorean realism as one of its real advantages. On his view, Mackie was right to be concerned with the thought that things out in the world might have the ability to give us reasons – that things in the natural domain could be inherently normative and motivationally efficacious. While other views require that there exist queer properties in the world, Scanlon’s contractualism required only that people had the desire to act justifiably.
In the 1980s, when Scanlon published CU, it was less popular than it is now to discuss normativity in the language of reasons. But what Scanlon himself later came to see is that his early view amounted to an assumption about reasons: namely, he assumed that desires could provide reasons to act. Scanlon’s early contractualism, then, fits neatly into a well-established class of metanormative theories. Views in this class attempt to explain normativity in terms of the passions, or more particularly, desires, and so are often called ‘neo-Humean’ or just ‘Humean’ accounts. Thus, by explaining one’s reason to be moral by reference to the desire to act justifiably, Scanlon formulated a Humean, Kantian contractualism.

Even according to Scanlon, then, contractualism need not be paired with primitivism. On the view articulated in CU, there is something informative and non-circular one can say about at least some reasons – and importantly, about the reason to act morally: namely, that it is due to a desire to be able to justify oneself on terms it would be unreasonable of others to reject. Not yet having accepted primitivism, then, Scanlon could have opted to explain ‘reasonable rejectability’ in a similarly Humean fashion, on which reasons to reject certain principles were also understood to be generated by desires. And had he done this, then any problems with Scanlon’s contractualism that are due to his adoption of primitivism (including not only the issues raised earlier in this chapter, but also any lingering metaphysical and epistemological worries one might have) could have been avoided. And to that degree, the view would have been stronger.

---

21 This admission comes in the Introduction of (1998, 6-7), in which he explains his move away from this view.
22 For a brief discussion of how the view is and is not Kantian, see Chapter 1 (f. 8).
23 And there is some evidence that this is what Scanlon was in fact trying to do in CU. See further his discussion of contractualism’s advantage over utilitarianism concerning its ability to explain the value of welfare (1982, 278).
Two problems remain, however. The first is that Scanlon abandoned his Humean foundation, presumably for good reason. And indeed, as I will preview in the final section of this chapter, retaining a Humean theory of reasons has theoretical costs. Before we get to that, however, there is the more obvious issue, I think, that Scanlon’s Humean account of moral motivation simply wasn’t a very plausible one. Although his updated account of the motivational basis of morality employs primitivism, *qua* theory of moral motivation it is far superior. Fortunately, however, the two views – Humeanism about reasons and Scanlon’s new and improved account of moral motivation – can be easily joined. So joining them is the goal of the following section.

2.4 How to Answer the ‘Whether’ Question

By looking at Scanlon’s early and later accounts of moral motivation, we can draw a general lesson about employing a theory of reasons in one’s moral view. Since the challenge of the WM question is to provide both a justificatory and motivational story concerning why one ought to do the right thing, it is natural to answer the question by reference to reasons. Because the motivational basis of morality must also contain a normative element, we are looking both for the reason one has to act rightly, as well as the reason that *does in fact move* morally sensitive individuals. On Scanlon’s updated view, there simply are reasons related to morality, and so the question is which of these existing reasons provides a plausible motivational explanation. A similar question might be asked by a value-based reasons theorist, like Derek Parfit (2001), who could canvass the morally-relevant values and ask which one gives rise to the reason paradigmatically acted on by morally sensitive agents. Scanlon’s initial articulation of his view, however, provides a very different alternative: on this Humean methodology, Scanlon looks at
the set of considerations that are relevant to justifiability, picks out one as the relevant consideration for morally sensitive agents, and then explains its status as a reason by appeal to the fact that agents desire that good. The complete answer to the WM question, then, on Scanlon’s early view, is that one has reason to act morally because she desires to be able to justify herself to others on terms they could accept.

The problem with that particular answer, however seductive it might be to philosophers (who spend their lives in an effort to justify themselves), is that it is not clear how widespread such a desire really is. Scanlon claims that it is both widely held and quite powerful (1982, 277), but we might wonder if such beliefs are more hopeful than realistic. People seem to have a fantastically easy time living lives that they could not justify, and even if they were simply lying to themselves, it is easy to imagine that many people simply do not care about so living. But surely, one might think, everyone (or at least, nearly everyone) has a reason to act morally, and so a desire-based moral reason that is not widely-shared seems to miss an intuitive data point. And even if most people did, in fact, desire to justify themselves, it seems likely that for at least some people, such a desire could be too easily diminished and maybe even eventually abandoned. This is a problem for the view because, if the explanation of the reason to be moral were by reference to a desire that could easily be so manipulated, then it would turn out that many people could ‘opt out’ of having a reason to act morally.

If we borrow some of the analytical tools from Scanlon’s updated views on moral motivation, we might remind ourselves that half of what we want from our view is that it pick out a consideration that is ‘recognizably good’. In the previous chapter, I referred to the recognizability of the good with the language of ‘how wide we cast the motivational net’.
Although it is unlikely that there is a single consideration valued by all agents (that is: we likely cannot find a shared basis for reasons with the psychopath or unrepentant evildoer), we do want to identify a reason to be moral that is as close to universally valued as possible. The goal, I take it, is to find a reason that is so central to most of our lives that it is difficult for us to believe (even if it’s occasionally true) that some people might simply be uninterested. The reason generated by our desire to justify ourselves, however, is not like that; it is all too easy to imagine telling someone that she has a reason to phi because otherwise she won’t be able to justify herself, and to have the agent respond, “Huh. Yeah, I don’t think I really want to justify myself all that much.” Note that there are many possible reasons for someone saying that, including the less-interesting ones like: that she is lying, both to us and to herself. For our purposes, though, the important point is that it’s not unimaginable that she could be telling the truth – that on reflection, either she doesn’t really want to be able to justify herself, or that she wants to, but not very much, and so decides that she can eradicate herself of that desire. The ease of imagining such a case casts doubt on whether we have really found a reason to be moral, the goodness and weight of which is widely recognized and difficult to ignore.

Utilizing the Humean grounding for contractualism from CU does not, however, require that we adopt the particular moral motivational account there developed. Indeed, when moving from the ‘which’ question in moral motivation (which reason, among some set, is the operative moral reason) to the ‘whether’ question (whether there is a moral reason), all that is required is that we take the consideration singled out to do the moral motivational work, and then ask in virtue of what such consideration is a reason. The primitivist’s answer is simply that it is. The value-based theorist will say similarly that the consideration is a reason in virtue of some
objective value – so, for instance, the value of living with others on justifiable terms gives everyone a reason to so live. The Humean, however, will claim that some desire is ‘well-served’ by acting in the relevant way, which is why there is a reason to so act. According to Mark Schroeder, this relation is cashed out in the language of ‘promotion’ (2007); so one’s desire, say, to live with others on terms they could accept (if such a desire exists) explains the reason to act justifiably, since so acting promotes the object of desire.

One way of bolstering Scanlon’s early account of moral motivation, while embracing his Humean formulation, then, is to look for another candidate consideration that could serve as the foundation for a more plausible Humean reason. However, the work done in Chapter 1, I contend, has already delivered this. According to the arguments presented there, the good of intimate relationships explains the reason to live with others in mutual recognition. My claim, then, is that the having of (at least some) intimate relationships is the object of widespread, powerful desire, and so would serve as a much more plausible Humean grounding for Scanlon’s view. While it is unclear how many people really do desire to be able to justify themselves, and how ‘escapable’ such a desire is, it does not seem similarly questionable how many people desire intimacy (or how easy it would be for them to abandon such a desire). Since the arguments of Chapter 1 claim to show that having any intimate relationships is promoted by being a ‘mutual recognizer’ (committing to living with others justifiably), all that is required to generate the relevant reason is the very weak (and seemingly nearly universal) desire for some intimacy. So long as I desire intimacy at all, this desire generates a reason to be a mutual recognizer, which, according to Scanlon, provides a reason to act justifiably. The Humean contractualist, then, can adopt the sophisticated answer to the ‘which’ question of moral motivation articulated in Chapter
I, and then explain the existence of the reason to be moral by reference to the *desire* to be in intimate relationships (thereby answering the ‘whether’ question as well).

As the comments made here are thus far only schematic, I will, in the next section, provide a more detailed sketch of how a Humean contractualism may actually be articulated.

2.5 A Humean Contractualism

At the conclusion of the previous chapter, I argued that the best solution to ‘Prichard’s Dilemma’ was to stop looking for a single reason to act rightly, and instead to look for a pair of reasons that together explain the competing *desiderata* of a moral motivational account. My substantive suggestion, then, was that a good candidate for a contractualist view of the motivational basis of morality is as follows: an act is right just in case it is justifiable; one’s reason to act justifiably, then, is the fact that so acting is part of being a mutual recognizer, and one’s reason to be a mutual recognizer is that failing to do so reliably blocks one from the having of intimate relationships. The grounding reason for acting rightly, then, is that doing so reliably promotes intimacy, but the middle term in the explanatory story – the reason to be a mutual recognizer – provides an account of the motivating reason that morally sensitive individuals are likely (at least sometimes) to act on.²⁴

²⁴ Further, being a mutual recognizer or acting justifiably could also be objects of desire, in which case there would be no need to reference a desire for intimacy in the explanatory story; indeed, one might (although this would almost certainly be a non-standard case), desire strongly to act justifiably, but not care one whit about intimacy; in this case too, the agent would have a reason to act justifiably. A nice implication of the complex moral motivational story above is that it has the capability of explaining the reason to be moral for those who do not have this more direct desire to act justifiably, but it does not require so utilizing a desire for intimacy. Many of us, then, will have multiple reasons to act rightly, as we desire intimacy *and* we desire to live with others on justifiable terms (or more atomically, to be able to justify our individual actions). Thus, although I have, for ease, discussed the explanatory story as if the desire for intimacy
The advantage of explaining one’s moral reason in terms of intimacy, I argued, is that such a good is more obviously and recognizably valuable, and so ‘casts the motivational net wider’. In switching to a Humean grounding for Scanlon’s theory, the good of intimacy will play a similar role. The concern I raised for Scanlon’s own Humean account was that the desire to be able to justify oneself seemed unlikely to be all that widely-held, and too easy to abandon. However, switching the focus to an explanatory story that features the having of intimate relationships as the grounding explanatory good helps to address this worry. So on the Humean view, rather than assuming that there is a reason to have intimate relationships – or that intimacy is objectively valuable – we will simply point to intimacy as the target of a widely held, powerful desire. That desire, then, grounds a reason to act in ways that promote its object. And, thanks to the work done in Chapter 1, we know that one way to promote the having of intimate relationships is to become a mutual recognizer, and part of being a mutual recognizer is acting justifiably. So the desire to be in intimate relationships, on a Humean account, explains why one has reason to act justifiably.

As we are moving beyond schematic explanation at this point, let me explicitly set out the initial pieces of the view I am developing. The Humean Theory of Reasons (HTR) has come a long way since Scanlon first introduced his view, and much of that advance is due to work done by Mark Schroeder (2007). Although I will explain in the following chapter that my own Humean commitments prevent me from simply adopting Schroeder’s view in total, I take it that always explains the reason to be in relationships of mutual recognition, this need not be the explanatory structure in every case. Some agents may have a more direct justification of justifiable action, with or without the additional reason grounded in a desire for intimacy.
his basic Humean analysis of what it is to be a reason is the best account going, and so I will take on an analysis very similar to his. The view that I will call the Humean, Contractualist Account of Moral Reasons (HCA), then, is constituted by the marriage of what I take to be the best versions of moral contractualism and metanormative Humeanism.

As I begin to build HCA, it is worth recalling from the Introduction that one goal of this dissertation is to argue in a philosophically modest way, attempting to generate substantive moral conclusions from relatively humble beginnings. As a result, I began this investigation by suggesting one way in which contractualism could be read, by one who wishes to reject Scanlon’s substantive view, as a merely formal theory that posits a biconditional between the set of justifiable acts and the set of permissible acts. As I now set myself to undergird the moral contractualist view with a particular view of reasons, it is worth asking: how modest is such a move? In one sense, joining Scanlon’s contractualism with HTR is clearly a move in the direction of modesty, as one of the great virtues of metanormative Humeanism is its ontological simplicity. As I like to say about it: if you believe in facts and desires, then you can accept (on ontological grounds) the Humean view of reasons. However, by positing a particular explanation of all reasons – and further, by positing a reduction, as I will just below – I make a move that is less modest than it needs to be. After all, I could have just claimed that some desires explain the existence of some reasons, and that would not commit me to a rather immodest claim about what it is to be a reason.

In response, I want to say something similar to what I said in justification of starting with contractualism: for much of the dissertation, one can read my invocation of HTR in the modest way just suggested. That is, just as one can, for the most part, read my use of ‘contractualism’ as
implying only formal contractualism, one may also (for the most part) read my use of HTR as implying not the strong, reductionist view, but rather the weaker claim that some reasons are explained by desires. In both the case of contractualism and that of Humeanism, one can read my arguments in this more modest way and still generate and evaluate the proposal that HCA explains the reason to be moral. This, I think, would still be a success. However, as I have already previewed in the Introduction: in the final chapters of this dissertation, I will ask whether HCA might be more than merely surprisingly successful for a modest view; in the end, I will ask whether HCA might be at least initially defensible, such that it ought to be taken seriously as a candidate view of moral reasons. For this reason, when it comes time to defend HCA, I will need a substantive version of the view to work with. And, to the extent that utilizing the modest version of HTR (that only some reasons are explainable by desires) is a substantive view of reasons, it is less modest for its openness. Indeed, such a position has its defenders, under the name of Hybridism about reasons, and the very goal of being a Hybridist is, among its defenders, to explain reasons that seem psychologically-based as psychologically-based, while explaining reasons that seem ‘given’ (or primitive) as given. But since my very interest in moving towards a Humean grounding for contractualism is precisely in order to get away from Scanlon’s primitivism, I will not here leave the door open for a Hybrid view. Instead, as with my introduction of formal contractualism, I simply invite any reader who finds HTR to

---

Ruth Chang is one of the few defenders of the Hybridist position (2009; 2013), although Nico Kolodny has also, recently, advanced a Hybridist position in (2011). In a nice summary look at the critical reaction to Scanlon’s philosophical contributions, Sarah Paul and Jennifer Morton (2014) make the interesting claim that many problems for Scanlon recommend that he adopt a Hybridist view. Scanlon himself, however, has publicly rejected the need for and defensibility of any view that explains reasons in two different ways (2004).
incredible to entertain to read my use of the view, for as long as possible, as the modest claim that some reasons can be explained in the way HTR claims.

Most simply put, then, the view I advance here is the product of joining Scanlon’s moral theory with Schroeder’s analysis of reasons, combined with the account of moral motivation developed in these first two chapters. The resulting view is constituted by three claims:26

**The Moral Theory:** an act is impermissible if and only if it would be disallowed by any set of principles that are acceptable to all; summarily, an act is impermissible if it is unjustifiable.

**The Theory of Reasons:** for some fact R to be a reason for agent A to phi is for A to have some desire, the content of which is p, and for the truth of R (at least partially) to explain why A’s phi-ing promotes p.

As is likely obvious, **The Moral Theory** is simply an articulation of Scanlon’s view (1982; 1998), and **The Theory of Reasons** is a very close adaptation of Schroeder’s (2007). The final piece, then, is:

---

26 The account of moral motivation, however, plays a different role than the contractualist and Humean elements. **The Moral Theory** and **The Theory of Reasons** are the two central analyses that constitute the Humean Contractualist position. **The Motivational Basis of Morality** is not an analysis that partially constitutes the view, and in fact follows from the Humean and contractualist claims (assuming the success of the arguments I have made along the way). So the simplest version of my view would in fact just include the moral and metanormative accounts; I include the motivational claim, however, as I take its vindication to be a central part of what makes the total view plausible. If I were convinced that **The Motivational Basis of Morality** were false, I would be much less confident in the Humean Contractualist position.
The Motivational Basis of Morality: agent A has a reason to be moral, in the way characterized by The Moral Theory, if and because A desires to have intimate relationships.

There are two things to note about The Motivational Basis of Morality, both due to the connective ‘if’. The first is that this is not a biconditional: although having a desire for intimate relationships is sufficient for having a reason to act justifiably, it is not necessary. Since the entire Humean Contractualist view is constituted by these three claims, it follows that any other desire that would be promoted by acting justifiably also gives one a reason to so act. The focus on the particular desire for intimacy, then, is that the arguments of the previous chapter make it a particularly good candidate for explaining both the motivational and justificatory elements of moral motivation. While Scanlon’s (1982) ‘desire to be able to justify oneself’ will also, in fact, ground a reason to act justifiably on this account, it is not the best explanation of the competing moral motivational desiderata.27

Second, The Motivational Basis of Morality is clearly hypothetical. If one has a desire for intimacy, then one has a reason to be moral. Now since this is only a sufficient condition, one might have other grounds for having a reason to be moral, but the view doesn’t guarantee this outcome. For this reason, HCA cannot guarantee that everyone has a reason to be moral; in other words, HCA fails to guarantee that morality is universally normative. This is not to say that my account should embrace the implication that whether agents have reason to be moral will vary widely; indeed, the very concern to find a widely held and powerful desire to ground the

---

27 See further f. 24, above.

75
moral reason will help to explain why the moral reason may seem universal and inescapable, even if it is not. However, my Kantian and realist opponents will be quick to point out that the hypothetical nature of the view is a cost – a cost that Scanlon escaped by becoming a primitivist. While I will do my best in the remainder of the chapter and dissertation to allay that concern somewhat, I admit up front that the failure to capture universal normativity is simply a cost of Humean Contractualism.

What, then, was gained for this price? In the language introduced earlier: HCA provides a genuine explanatory payoff that justifies the foray into metaethics. While a pure, first-order contractualism can be motivated by an urgent sense of practical demands, the task of providing a metanormative foundation for some constructivist view is undertaken out of an interest in having a full explanation. My criticism of Scanlon’s view above is that he takes on theoretical burdens in order to embed his contractualism in a broader theory of reasons, but that this move frustrates rather than elucidates. The primitivism that he imports doesn’t really provide any additional explanatory power, and it seems capable of replacing the first order view it justifies.

A Humean Contractualism, on the other hand, provides the complete explanatory story. Is there a reason to be moral? Yes (for most of us), because we desire states of affairs that are promoted or enabled by acting morally (justifiably). On one way to understand the Humean position (Schroeder 2007), it is a reduction: the reason to be moral is constituted by some natural fact, given a particular background condition (the relevant desire). One way, then, to improve on

---

28 I take it that this is the move made by Rawls, for the political domain, in (1971).
29 In her discussion of ‘The Normative Question’, Korsgaard similarly claims that what is frustrating about (‘substantive’) realist views like Scanlon’s is that they feel ‘stipulative’: we ask why, for instance, one has reason to act morally, and the realist responds, ‘We do’ (1996).
Scanlon’s contractualism is to pair it with a reductive account of the reason to be moral, rather than a realist one.

Consider another way to improve on Scanlon’s view. One might, with Sharon Street, hold that a normative constructivist view such as Scanlon’s ought to be constructivist at the metaethical level as well. On Street’s view, such a ‘thoroughgoing constructivism’ reapplies the constructivist methodology to explain not just moral reasons, but all practical reasons, with the result that all reasons for action are generated from inside the perspective of an already-situated, already-valuing agent.\(^\text{30}\) A thoroughgoing constructivist then, or a ‘constructivist about reasons’, holds that something is a reason for an agent because the judgment by the agent that it is a reason withstands scrutiny from her other normative perspectives (Street 2008; 2010; 2012).\(^\text{31}\)

It is unclear how much difference there is between a reductive view like my own and a constructivist view like Street’s. While the reductionist looks to be doing metaphysics, the Humean reductionist and the metaethical constructivist are in agreement about the essential point: the truth-makers for normative facts come from within the perspective of a valuing agent. Further, Street is a ‘Humean Constructivist,’ because she denies that universal moral reasons can be guaranteed by the formal constructivist machinery; if reasons are generated by the values of

\(^{30}\) In fact, Street believes that all reasons – including epistemic ones – can be explained in the same way, although this is a bigger claim than she or I need to make in this context. She admits this in (2010, f. 4) and begins the project of making good on this more ambitious claim in (2011).

\(^{31}\) There are obviously many questions at this point concerning how such a view ‘bottoms out’, since normativity is analyzed in terms of judgments of normativity. Although Street has extensive discussion of how such a view works (2008; 2010; 2012), I will not go into details here, since I will not be adopting the view. My goal here is only to highlight that one who is moved by the challenge I make to Scanlon has an alternative to reductionism; one could, with Street, attempt a constructivist grounding to Scanlon’s constructivist normative theory.
agents, then idiosyncratic agents will have idiosyncratic reasons, and all overlap between agents’ reasons will be contingent and coincidental. This, Street takes it, is Humean because an agent may very well have reason to prefer the destruction of the entire world to the pricking of her little finger (Street 2010).

Street does, however, have a Kantian aspect to her view as well: although Humean when it comes to the question of whether universal morality is likely to emerge, she is Kantian when it comes to the question of what kind of attitude generates the normative facts. On Street’s view, one has a reason when her judgment that she has a reason withstands scrutiny from her other normative perspectives. The reasons that one has thus results from what one ‘legislates’ from the practical perspective, not from ‘mere desires’ (Street 2012).

A reductionist view like The Theory of Reasons and a constructivist view like Street’s are thus similar in their requirement that normative facts be grounded from within the perspective of an already-situated agent, but differ structurally in whether a property analysis of the normative phenomenon is being provided (the reductionist provides one while the constructivist doesn’t) and in content concerning what attitude or perspective ultimately grounds normative truth (the Humean reductionist claims that the relevant attitude is desire, while the constructivist holds that it is ‘judging something to be a reason’). While the structural difference seems essential to each position, the content difference seems contingent: Street could have made

---

32 This may only be ‘Humean’ in the broad sense of ‘inspired by Hume’, as it may not, in fact, track very closely to Hume’s own views about normativity. I suspect that, in fact, this is the case in most contemporary uses of the language of Humeanism, but nothing substantive hinges on the point. As I say in the main text: for Street, the language of ‘Humeanism’ and ‘Kantianism’ simply allow for multiple, helpful distinctions that bear some relation to the historical positions of the views’ namesakes.
her view a Humean *Humean* constructivism (rather than a Humean *Kantian* one), according to which no substantive moral truth is guaranteed by the constructivist process (the first sense of Humean) and whether an agent has a reason is determined by whether her *desire, or passion* forms part of a coherent perspective.

At this point, I am unsure how important the structural distinction between a Humean reductionism and a Humean (Humean) constructivism is; I am also unsure whether one view is likely to be stronger than the other. For the purposes of this dissertation, I will adopt the reductionist model, for the simple reason that it seems more accurate to me. On reflection, when attempting to explain the reasons that we have, it simply *seems* more accurate to say that we are asking an ontological question, and in providing a Humean answer, that we are giving a genuine property-analysis. However, the commonality between the reductionist and the constructivist (the rejection of realism, and the insistence on perspective-dependent grounding for normativity) seems far more important than this mild preference, and so I take it that I could be convinced to switch teams if the constructivist could be shown to have the stronger position. For this reason, I do not want to make an enemy of the constructivist; but rather, I want to highlight the way in which Street, for instance, and I are engaged in the same project of trying to provide a genuinely explanatory grounding for restricted constructivist views like Scanlon’s.

2.6 A ‘Back-Up’ Theory of Moral Reasons

A Humean Contractualist Account of Moral Reasons has many virtues. It promises a complete explanation of moral reasons, with very little theoretical buy-in. It is ontologically simple and explanatorily powerful. If you believe in facts and desires, then my arguments claim to be able to generate a plausible theory of moral reasons that you can accept. On this view,
facts are reasons if and because they stand in a certain relation to desires; people largely have a
strong desire for intimacy; a desire for intimacy generates a reason to be a mutual recognizer;
mutual recognizers act justifiably; and justifiable action is permissible action. On my Humean
Contractualist view, therefore, each person’s desire for intimacy (if they have it) explains a
reason for that person to act justifiably.

I did admit, however, that there is a cost, and it will have been obvious to my opponents.
I concluded above that each person’s desire for intimacy if they have it explains that person’s
reason to act justifiably. But, naturally, we shall want to know: and what if a person does not
desire intimacy?

The answer to this question is complex, but at the end, I must admit what my opponent
suspects: if a person does not desire intimacy, then that person lacks the reason in question to be
moral. And having admitted this, some will conclude that HCA therefore fails. The universal
normativity of morality is a non-negotiable desideratum for a moral account, one might say; or it
is a conceptual truth about morality that it has ‘clout’,\textsuperscript{33} which is just to say that deontic facts
always provide reasons. However, I want to beg for patience at this point, as I did warn that my
answer to the challenge is complex.

The first thing to remember is that \textbf{The Motivational Basis of Morality} is merely a
sufficient condition for having a reason to be moral; other desires will generate a reason to act
justifiably as well, and so the absence of a desire for intimacy only guarantees that one does not
have \textit{that particular} reason to so act. People will often have other, idiosyncratic reasons to act

\textsuperscript{33} See Joyce (2006).
justifiably as well. Justifiable action, for instance, is often the kind of action that keeps one from being ostracized, and so the desire not to be kept out of *all sorts of relations* (not merely intimate ones) will also sometimes generate the reason to be moral. Or even less admirable desires, such as the desire (either instrumental or basic) to be seen as a virtuous person, will also often generate the moral reason. Even the very abstract, common, standing desire *to succeed* is a great candidate for a desire that may generate the moral reason, as success is often (although certainly not always) promoted or enabled by justifiable action. While I certainly do not here intend that HCA can, after all, guarantee universal normativity for morality, I do want to show the myriad ways in which particular, worrisome agents can still be seen to have reason to act justifiably.34 I take it that such observations make the class of individuals with no reason to be moral even smaller than it would be just given *The Motivational Basis of Morality*.

Since I have not forestalled the very possibility of there being such individuals, however, many will be unconvinced by the prospects for HCA. In the following chapters, I will in fact attempt to make the strongest case for accepting HCA, as I think the theoretical virtues outweigh

34 One may wonder why, given all of these alternative possible reasons to act rightly, I have singled out the desire for intimacy. The answer to this question is essentially the argument of Chapter 1, but it is worth summarizing again for the present context. Because I am convinced by the challenge of Prichard’s Dilemma, I believe that ‘the reason to be moral’ is likely *complex*, consisting of two reasons – one that satisfies the motivational aspect (what those who care about morality are moved by) and one that satisfies the justificatory aspect (why everyone has reason to care about morality). The desire for intimacy is plausibly unique in that it generates a weighty reason for a very large class of people *via* the ‘motivational’ reason to be moral – that is, the reason to be a mutual recognizer, or a ‘justifier’. So while I am happy to lean on the existence of very many other reasons to act in permissible ways in order to comfort us that not very many people completely escape having a reason to be moral, I think it is unlikely that any other candidates share this unique explanatory relationship to justifiability. However, I should note that, were I convinced that there were others like this, I would happily fold them into my explanation, as it would not be a challenge to my view, but rather, a friendly addition.
the costs. In what follows, however, I want to return to the emphasis on philosophical ‘modesty’ first raised in the Introduction, and to suggest an alternative way to read the remainder of the dissertation.

To those who simply will not take seriously a view that fails to vindicate the universal normativity of morality: consider the arguments in the rest of this dissertation as being for a ‘Back-Up Theory of Morality’. I mean this literally. A ‘back-up’ refers to someone or something that is in position in case the primary candidate (for whatever function is being fulfilled) doesn’t work. While I hope that some others, like me, will consider HCA as a primary candidate view, I simply ask that others consider it as a back-up. Why do we need a back-up? Because other, more ambitious views are difficult to defend, and may be false. I, in fact, think that they probably are. But I can’t be sure, and neither can anyone else. So a helpful question to ask is the following: what would be true about morality, even if we could rely only on the most minimal theoretical assumptions? The arguments of the first two chapters have defended the claim that, even in that case, morality would look about like we think it does for a large group of people. Everyone who desires to have intimate relationships has a reason to act morally, and it seems a pretty good guess that most of the people we interact with have that desire.

In the remaining chapters, I will continue to fill in this position. Even as a ‘back-up’, the Humean Contractualist account thus far described is quite incomplete. Before asking how successful the view is, more details are needed – in particular concerning the weight of the reason to be moral (and so, for reasons in general) – and so filling in these details is the goal of Chapters 3 and 4. Providing such details will only make clearer the challenges I face, however, and so in chapters 5 and 6, I will raise and respond to the most threatening of these worries.
Finally, in the concluding chapter, I will come back to the idea of a back-up theory of morality. There, I will look at what HCA accomplishes, and what defects render it a ‘back-up’ view. While I take it that the exercise in working out the details of a back-up theory is valuable regardless of how convincing such a view ultimately is, I will suggest in closing that HCA ends up being so successful by any reasonable criteria that regardless of one’s antecedent commitments, its viability should make us feel a little better about the status of moral truth. Even if the worst-case scenario is the actual scenario, and all other, more ambitious views are false, a view that gets us very close to what we want out of an account of moral reasons is still likely to be true.
CHAPTER 3

In the previous two chapters, I have begun arguing for my own Humean Contractualist view of moral reasons, according to which an act is wrong if it is unjustifiable, and the reasons one has to act rightly are explained according to the Humean Theory of Reasons (HTR). However, as I admitted at the end of the preceding chapter, even supposing that I have made this marriage of views plausible, much remains to be explained. In particular, while I have argued that most people, on the Humean view, will have at least one reason to act according to the contractualist standard of permissibility, I have said nothing about how weighty that reason is. Does our reason to act rightly reliably outweigh our reasons to do otherwise? Intuitively, it seems that it should. The demands on a theory of moral motivation are not just that it provide an account of the reason to act rightly, but that it explain the (seeming) fact that this reason is the one that ought to be decisive; a fully satisfying moral motivational story will explain why we have the weightiest reason, or most reason, or perhaps even why we ought to act rightly.¹

Having adopted HTR, it might be supposed that I have already settled on a theory of weighting, as HTR is often thought to have a very natural theory of weighting at the ready. Recall that on the Humean view I have adopted, for R to be a reason for A to phi is, generally, for A to have some desire that p and for the fact of R to explain how phi-ing promotes p. So, for example, the fact that cool, clean water flows from my faucet is a reason for me to fill my glass from the tap because I have a desire to drink, and the fact that water flows from my faucet explains how filling my glass from there promotes my drinking. What is the ‘weight’, then, of

¹ The relation between these normative concepts will be discussed briefly in 3.1, and in f. 6.
that reason? Well, the answer seems to suggest itself: that depends on how strongly I want to
drink, and to what degree filling my glass from the tap promotes my drinking. And so it goes for
all reasons: the weight of a reason is doubly proportional, for the Humean – first, to the strength
of the grounding desire, and second, to the degree to which the action in question promotes that
desire. This view has been called, appropriately enough, proportionalism.

Proportionalism seems to have been the default Humean view, without careful
explication or defense, until quite recently.2 However, Schroeder, whose formulation of HTR I
borrowed and adapted, in fact argues that there is not any real pressure to accept proportionalism,
(2007a, 100; 2007b, 106), and that replacing it with a different view renders HTR much more
plausible (2007a, 97-102). The adherent of HTR thus should reject proportionalism in favor of
what Schroeder calls a ‘recursive analysis of the weight of reasons’, according to which some
reason’s having a weight is analyzed in terms of it being correct to place more weight on that
reason in deliberation (2007a, 137).

Critics who have pushed back on Schroeder’s arguments have largely been opponents of
HTR. David Enoch and Daan Evers, for instance, who argue carefully and convincingly that
Schroeder is wrong to reject proportionalism while embracing HTR, do so precisely because they
see HTR’s tight connection with proportionalism as one of the best reasons to reject the Humean

2 It is interesting to note that virtually everyone who writes on proportionalism seems to believe
this – that it is the default view – and yet I have not found a single author who cites past
Humeans as evidence for this claim. So, for instance, Schroeder (2007a), Enoch (2011), Sobel
(2009) and Dancy (2012) all make the claim (which I believe is true) that proportionalism seems
to be the standard theory of weighting for a Humean, but in none of these cases is this claim
followed by a citation. Perhaps this is evidence for how obvious the philosophical community
finds the marriage between HTR and proportionalism, as it does not even require precedence.
position (Evers 2009; Enoch 2011). And this, of course, is likely one of the reasons that Schroeder is motivated to abandon the proportionalist position.³

This dialectic has the interesting result that discussion of proportionalism – a view that everyone seems to agree is orthodox to the Humean position – is carried on entirely by critics of proportionalism. In this chapter, I hope to work towards remedying that situation. Building on some of the gestures and criticisms made by Enoch, Sobel and others, my goal here is to make the positive case for why a Humean should also be a proportionalist. And the answer, in short, is quite simple: because the view is natural and intuitive, and because there is no good reason for a Humean not to be a proportionalist. Slightly more carefully, the argument proceeds as follows: there is clear prima facie reason for the Humean to adopt proportionalism, and so one would have reason to resist this pressure only if there was either a better candidate view available, or if proportionalism failed seriously to satisfy some theoretical desideratum. Making the case for this initial pressure is the goal of section 3.2. I will then, in 3.3-3.6, go on to argue that even supposing Schroeder is able to motivate abandoning proportionalism, his own alternative view

³ Proportionalist Humeanism is supposed to be implausible because, as I will discuss further at the end of Chapter 4, it amounts to the claim that one has most reason to act in whatever way best promotes her desires. However, as Enoch (2011, 438-439) notes, whether it is Schroeder’s motivation or not, the implausibility (if it is implausible) of a proportionalist Humean view cannot be the primary justification for rejecting proportionalism. At a minimum, Schroeder must at least show that proportionalism is not the only plausible candidate theory of weighting from the perspective of a Humean. Otherwise, the argument would be clearly ad hoc, rejecting the most plausible candidate theory of weighting because not doing so is devastating to the overall view. Now, since I am a Humean, and about to argue in favor of proportionalism, obviously I disagree that adopting proportionalism is devastating to the overall Humean project. However, where Enoch and I are aligned against Schroeder is in our agreement concerning the options: We can either adopt a proportionalist version of HTR, or we can abandon HTR, but we are not justified in adopting HTR while rejecting proportionalism. I return to this point briefly in section 3.7.
(as well as a variant on his actual view) fails. As a result, the Humean must either resist Schroeder’s case for abandoning proportionalism, or come up with an alternative that avoids the pitfalls of both proportionalism and the Recursive View. In section 3.7, I argue that a Humean can resist Schroeder’s case against proportionalism, and so is left with no good reason to abandon the view, and good reason to endorse it.

3.1 Schroeder and the Challenge of Weighting

In the previous chapter, I adopted, as my Theory of Reasons, Schroeder’s version of Humeanism. As a reminder:

The Humean Theory of Reasons (HTR) is the view that for some fact R to be a reason for agent A to phi is for A to have some desire, the content of which is p, and for the truth of R (at least partially) to explain why A’s phi-ing promotes p.

In that initial argument, I made the case for HTR from the perspective of contractualism, and so did not make any mention of why one might be a Humean in general. However, in the following section I will suggest that the best reasons to believe HTR also provide reasons to believe proportionalism, and so I will here take a brief detour to note what exactly those reasons are; that is, I will here briefly make the case for HTR on its own merits.

There is a classical argument that infers HTR from the Humean Theory of Motivation and Reasons Internalism, but I agree with Schroeder that this is not the best reason (in fact, not even a good reason) for adopting HTR. Instead, the best reason for adopting HTR is its intuitive

---

4 The Classical Argument runs as follows:
(1) Existence Internalism About Reasons: If there is a reason for a person to do something, it must be possible to motivate her to do it for that reason.
(2) The Humean Theory of Motivation: Motivation requires desire.
plausibility for a wide range of every-day cases. Consider the one that Schroeder provides – the
case of two men, Bradley and Ronnie, who are considering going to a dance party (2007a, 1).
Ronnie likes to dance, while Bradley hates it. Our intuitions, Schroeder suggests, tell us that this
psychological information about Ronnie and Bradley explains the following normative facts:
Ronnie has a reason to go to the dance party (namely, because there will be dancing there), while
Bradley does not (in fact, that there will be dancing at the party is a reason for Bradley not to go).
Further, Schroeder eventually argues that the aspect of Ronnie’s and Bradley’s psychologies that
explain these normative facts are desires: Ronnie desires to dance, while Bradley does not.5

I take it that such examples can be multiplied ad nauseam, concerning thousands of
reasons and actions each day. Right now, facts about what I want provide me with reasons: to
keep writing; to take a break; to get some water; to pet my dog; to refill my coffee; to read more
literature on reasons; to prepare my lecture for this week’s class; and likely a hundred other
things. And for the Humean, the reference to an agent’s desires is precisely the right kind of
explanation for why some fact is a reason for that agent. That spending more prep-time on my

(3) Therefore, HTR: Having a reason requires having a desire.
Schroeder discusses in (2007a) several reasons for being skeptical of the Classical Argument, but
my own view is that the argument requires the defense of contentious theses – premises (1) and
(2) – and is not necessary; thus, we should avoid it if possible. And avoiding the argument is not
only possible, it is advisable, since I think the intuitive argument that I give in what follows is a
much better, more natural reason to believe HTR.
5 Although I will, here and elsewhere, use Schroeder’s language of desire, this is only for the
sake of simplicity. I find it implausible that desire is actually the relevant psychological state, as
Ronnie and Bradley seem to have reasons just in virtue of liking and disliking, having an interest
or disinterest, or even having some degree of concern for dancing. Thus, I will use desire as a
kind of placeholder, into which one could insert a variety of ‘pro-attitudes’.

88
lecture will make class go better is a reason for me to do so because I want to be an excellent teacher.

Understanding what we have a reason to do is not, however, all that we want from a metanormative account. Reasons are of interest because we want to know what we have more reason or most reason to do. But how do we get from claims about individual reasons to claims about having ‘more’ or ‘most reason’? We need an account of how reasons compare to one another, or what is often called ‘a theory of weighting’. In order to be a complete metanormative theory, HTR, like any other theory of reasons, needs an account of how we get from the observation that I have conflicting reasons to the conclusion that I have more reason to do one thing than the other. We want to know whether in continuing to sit here and type I have done as the reasons recommend.

---

6 In fact, we want to know what we ought to do, or what we are required to do. Schroeder bridges this final gap by analyzing ‘A ought to phi’ as ‘A has most reason to phi’. The equating of ‘most reason’ and ‘ought’ is typical not only of a normative reductionist like the Humean, but also of many others who want to analyze the normative in terms of reasons. Cf. Smith (1994), Parfit (2001; 2011), Scanlon (1998). Since this reduction is a stronger view than I must take, I will avoid it in this dissertation. Instead, I will assume only what seems fairly uncontroversial, which is that a theory of weighting must explain how a reason to do X can be ‘weightier’ than a reason to do Y such that the result is that agent A has more reason to do X than Y; further, if the reason for A to do X is weightier than the reason to do anything else, then A has most reason to do X.

Now, just to preview, one may wonder whether even such a relatively modest assumption allows for Razian-style ‘exclusionary reasons’, since these seem to run counter to the simple additive view of weight. In fact, the view I end up defending in the following chapters requires the vindication of a kind of Razian ‘second-order reasons’ that allows for the existence of reasons to consider or not to consider other reasons. This is a purely psychological view of second-order reasons, and so may fall short of what Raz ultimately defends under the heading of ‘exclusionary reasons’. This is discussed in more depth in Chapter 5.
In response to the demand for a theory of weighting, HTR has an obvious response: how weighty a reason one has to phi depends on the strength of the desire that grounds her reason to phi, as well as the degree to which phi-ing promotes that desire. That is, I have a reason to walk down the hall because I desire water, and that reason has more or less weight in proportion to just how much I want water, and the degree to which so walking promotes my drinking.\textsuperscript{7} Schroeder aptly names this view proportionalism (2007a, 98).

Proportionalism is a very natural theory for HTR to adopt. Schroeder, however, says that while he realizes that Humeans feel pressured to accept proportionalism, he thinks that this pressure is ‘illusory’ (2007a, 100; 2007b, 126). He goes on to argue that proportionalism fails to account for some considered normative intuitions, but I want to suggest that such a concern, even if valid, fails to show that the pressure to accept proportionalism is illusory. Tallying up costs of the view may provide pressure to abandon a view that one was pressured to accept, but that would not mean that the pressure was not there or real.

\textsuperscript{7} In general, I will here focus on the first proportionality, as that is the key property of the view. Presumably, all theories of weighting will employ something analogous to the second kind of proportionality, as there are many ways to promote a desired or valuable end, and some of them will be much more successful than others. Thus, even a value-based theory of reasons will want to make sense of why the reason to reliably promote some good is weightier than the reason potentially, merely possibly to promote that same good. If hydration is valuable (as seems plausible), then I have a reason both to fill my glass from the sink and to step outside and look up with my mouth open, waiting for it to rain. But surely my first reason is weightier than my second, since taking my water-glass to the faucet promotes the relevant value to a very high degree, whereas standing outside with my mouth open very weakly promotes the same value. The proportionality between the weight of a reason and the degree to which a given act promotes the relevant end (desire, value, etc.) is thus not unique to proportionalism. This point will, however, return briefly in the following section, as I think the directness of explaining this obvious feature counts in favor of proportionalism, while the fact that Schroeder’s view will have to explain it indirectly counts against his.
The goal of the next section, then, is to discharge my first argumentative burden in the chapter, by arguing that the positive case for a Humean’s accepting proportionalism is quite strong, and so there is genuine pressure to accept it.

3.2 The Genuine Pressure for Humeans to Accept Proportionalism

My case for the plausibility of proportionalism is quite simple. It seems that our reasons for believing a theory of weighting should not be completely divorced from our reasons for believing the theory of reasons itself. So methodologically, if we want to know what makes some reason weighty, we should look to why we think the reason is, in fact, a reason. As I mentioned earlier, I agree with Schroeder that the best reason for believing HTR is not based on any abstract argument (say, inference from a theory of motivation), but rather on the intuitive evidence from everyday normative experience. Ronnie and Bradley have reasons related to dancing, I have reasons related to my writing, teaching, and coffee breaks, and we all have reasons of similar kinds to do all sorts of things. My methodological question, then, is: do these informative cases also provide support for any one theory of weighting?

Unsurprisingly, I think they do. Take a fairly standard case for many of us – that of the reason to get coffee. Every day, several times a day, I think it is true that I have a reason to walk into my kitchen, to boil water, to grind fresh beans, and so on, and that my reason to do each of these things is explained by my desire for good coffee. Now if we want to discover what accounts for the weight of a reason, we should imagine the case in which my reason to do those things is very weighty, and the case in which the reason is not weighty at all, and then ask: what changed in each of these cases? When we imagine that my reason to prepare coffee is very strong, the most obvious explanation for this fact that I can see (assuming HTR) is that in this
case my desire for coffee is very powerful. Perhaps I have been reading about a fantastic new method for brewing, and it has begun to make my mouth water. Whatever the cause for the desire, given the view that desires explain reasons, what seems overwhelmingly plausible in this case of weighty reasons is that the grounding desire is strong. And so it is with the converse case of a weak reason. If it is true of me right now, as I sit in my writing chair, that I have very little reason to get coffee, then (again, assuming HTR) what seems plausible is that I don’t have any strong, coffee-related desires\(^8\) (or that I am not in a position to greatly promote any desire that I do have).\(^9\)

My contention is that this story is an incredibly natural one for the Humean to tell, and that it generalizes to nearly all of the cases that made HTR plausible in the first place. For any example such as Ronnie and Bradley’s, on which the Humean leaned to make the initial case for HTR, the above story about weighting will seem correct: Ronnie’s reason to go to the party goes up as his desire to dance goes up. Similarly, Enoch suggests that if proportionalism is false, it seems a bit too convenient that the weight of Ronnie’s reason seems to go up exactly, well, in proportion to the strength of his desire (2011, 439). I then want to strengthen this challenge by

---

\(^8\) This is different from the case in which I have a writing deadline in one hour, and so my reason to get coffee is simply outweighed by my reason to continue writing. In that case, my proportionalist intuitions tell me that I might really want coffee, but also really, really want to make my deadline, and so I may have more reason to continue sitting here despite having a weighty reason to get up and make coffee.

\(^9\) So, for instance, I may have very strong desires to improve the situation of coffee farmers in South America, who provide my fair-trade coffee; and it is, in fact, the case that drinking more coffee right now will increase my need for more coffee, and so increase the business that I give those farmers. However, the extra cup that I could drink right now simply doesn’t promote that desire to a great degree, and so, in accordance with the second proportionality mentioned in footnote 7, I might still have a weak reason to get coffee, despite this particular, coffee- (and justice-) related desire being quite strong.
noting that it is also convenient that the strength of one’s reason seems to go up, well, *in proportion* to the degree that acting in the recommended way promotes the relevant desire. Proportionalism directly explains both of these features, and in a very natural way.

Schroeder’s response to this kind of argument is to claim that the proportionalist ‘over-generalizes’ from the case of Ronnie and Bradley, in that “we don’t have to hold that all reasons are like Ronnie’s in being proportional, in order to accept that all reasons are explained in the same way as Ronnie’s” (2007a, 99). It seems to me, however, that proportionalism generalizes to exactly the right degree: I simply cannot think of a motivating case for HTR that is not also a motivating case for proportionalism. This is not to say that *every* example concerning a reason is likely to motivate proportionalism. Schroeder and many others see moral reasons as the obvious counter-example to proportionalism, as the weight of moral reasons do not seem to diminish as the strength of the grounding desires weaken. However, moral reasons, I think (and opponents of HTR tend to agree), also never seemed the most plausible candidates for a Humean explanation to begin with. That is: yes, proportionalism seems less plausible in the case of moral reasons, but that’s because *Humeanism* seems less plausible in the case of moral reasons. But all of the cases that seem to make HTR plausible – cases like dance parties and coffee – also seem to make proportionalism plausible.

Perhaps that was too fast. Aren’t there non-moral reasons that seem to be counterexamples to proportionalism? Schroeder offers the case of his Aunt Margaret, who wants more than anything to reconstruct the scene depicted on page 78 of the November 2001 *Martha Stewart Living* catalogue on Mars (2007a, 84; 143). Now such a case was, Schroeder admitted, originally a counter-example to Humeanism generally, as it seems to imply that Aunt Margaret
has a reason to build a Mars-bound spacecraft (since no one is going to give her one!); however he argues that it is not implausible that Aunt Margaret has some reason to build such a craft, so long as it can be shown that she has very little reason. Schroeder’s claim, then, is that the view he ultimately defends is able to generate this conclusion, while proportionalism is not. After all, we can imagine that Aunt Margaret wants her recreation on Mars more than anything, and obviously building the spacecraft is necessary to the recreation.

Proportionalism, however, has no problem with Aunt Margaret’s case, as the strength of the desire is only one determinant of the weight of one’s reason. The other, recall, is the degree to which the recommended action promotes the desired end. But now we can ask: to what degree does, say, gathering space-craft material promote Aunt Margaret’s end? The answer: practically not at all, because Aunt Margaret simply cannot build a Mars-bound spacecraft. Her desire is crazy not just because the Mars recreation is worthless, but because Aunt Margaret will never be able to accomplish her goal, regardless of what actions she takes. Of course, it is a difficult question as to precisely how the two forms of proportionality intersect; does the very slight, next-to-nothing promotion of Aunt Margaret’s strongest desire generate a weightier reason than, say, perfectly promoting a weak, fleeting desire for some particular food? Answering such questions would require a more thorough account than I will provide here, but it is also unnecessary. What is important is that a virtually non-existent promotion of some desire, no matter how strong, will not yield a very weighty desire, and so proportionalism can handle the case of Aunt Margaret.

Of course, I have assumed that Aunt Margaret is not a Mars-Spacecraft engineer with unlimited funds, and so one might ask: what if she was? What if my assumption that Aunt
Margaret could not promote her strange desire was unwarranted? In such a case, a proportionalist Humean view would have to admit that, if collecting the materials to build her spacecraft significantly promoted one of her strongest desires, then she has weighty reason to do so. However, I also no longer find the case all that counter-intuitive. Aunt Margaret, under this description, falls into the category of what Sharon Street calls ‘Ideally Coherent Eccentrics’ (ICEs), along with moral monsters like the ideally coherent Caligula, and strange folks like Rawls’ blade of grass counter (Street 2009). ICEs do seem to be counterexamples not just to proportionalism, but to a complete proportionalist Humean view, and so I will deal with them at length in Chapter 6. For now, I will simply say that ICEs are very strange creatures, and make up a very small class of actual and possible agents. As a result, I think that Schroeder’s claim that proportionalism ‘overgeneralizes’ is much too strong; proportionalism is a view that naturally and directly explains virtually all cases that were motivating cases for Humeanism, but which has a strange implication for a small number of (likely non-actual) cases. This leaves work for proportionalism to do (again, see Ch. 6), but does not seem to show that proportionalism is intuitively weaker than it looks.

If I am right about the intuitiveness of proportionalism, then the very methodology that recommended HTR in the first place (namely, the intuitions from everyday cases) recommends proportionalism as the correct theory of weighting for HTR to adopt. And this, I think, should be unsurprising. For HTR is the view that reasons depend for their existence on the desires of particular agents. So what gives a reason more or less normative force? Well, the same thing
that makes it normative in the first place: the desire of some particular agent. Thus, there is a good reason for proportionalism to seem intuitive (the role that desire plays in explaining normativity), and the very methodology that Schroeder agrees gets us to HTR seems nearly univocally to recommend proportionalism. While this does not entail that we should ultimately accept proportionalism as the correct Humean theory of weighting, these two considerations certainly seem sufficient to lend the view *prima facie* plausibility, and so to suggest that there is genuine pressure for the Humean to accept proportionalism. Of course, as I said, genuine pressure is sometimes best resisted. Ought the Humean to resist proportionalism? It is to this question that I now turn.

3.3 ‘The Attractive View’ of Weighting

So far, it appears that proportionalism is the most natural theory of weighting for HTR. As mentioned above (f. 3), both Schroeder and critics of HTR seem to think that HTR’s adopting proportionalism is devastating to the overall Humean position, but this worry cannot be the reason for abandoning proportionalism if it is, indeed, the best account. Schroeder and Humeans who agree with him may *hope* that there is reason to abandon proportionalism, so as to save HTR; but there must be *something* wrong with the view in order to motivate abandoning it. So what is wrong with proportionalism?

The main problem with proportionalism, Schroeder seems to think, is that it fails to respect the apparent fact that “the weight of reasons is a normative matter” (2007a, 100). What he means by this is that to say some reason is weighty is not descriptive in the way it is to say

---

10 Thanks to Mark Murphy for this way of summarizing the intuition.
that some person weighs 200 lbs; rather, it is normative in the way it is to say that someone is admirable. To say that someone is admirable is to say that there are reasons to admire that person, and the same sort of claim must be true of a reason’s weight, Schroeder thinks. So just as it is correct to admire someone that is admirable, it seems that there must be correctness conditions relevant to a reason’s weight as well. It is thus a ‘truism’, Schroeder claims, that it is always correct to place more weight on a reason that is weightier (2007a, 101).

Proportionalism is ‘ill-suited’ to account for this truism, and so the correct theory of weighting will likely take a very different form (2007a, 101; 129-130). In particular, the correct view must explain how the weight of a reason admits of a normative explanation, and how such an explanation is consistent with HTR. And on the face of it, this will be a challenge, as a thing’s being ‘correct’ to do seems plausibly to entail that there is a reason to do that thing. In fact, in line with his general reductionism of the normative to reasons, Schroeder claims that for something to be correct just is for it to be the case that the reasons in favor of doing that thing outweigh the reasons in favor of not doing it (2007a, 134). But Schroeder’s ‘truism’ about weighting holds that it is always correct to place more weight on a weightier reason, and so it seems that one always has a reason to place more weight on a weightier reason. The ‘truism’ concerning the normativity of weighting, then, requires a ‘recursive’ explanation of weight.

Although I will eventually argue that proportionalism can respond effectively to the challenge of Schroeder’s ‘truism’, I admit the initial power of the suggestion. It is thus important, I think, to investigate closely whether Schroeder is able to develop an alternative to proportionalism that successfully explains the normativity of a reason’s weight. I will argue in the coming sections that Schroeder’s view fails. With no other candidate Humean theories of
weighting in sight, this means that either we must abandon HTR, or defend proportionalism against the charge that it can’t make sense of weight’s normativity. In the closing sections, I will argue for the latter strategy by showing that we can capture what seems correct about the ‘truism’ in a way that does not threaten proportionalism.

3.4 A Recursive Account of Weighting

As previewed above, there is something very strange about any position that tries to make sense of Schroeder’s ‘truism’ – namely, it adds a layer to our metanormative story. We now have correctness conditions on an activity of ‘weighting’ reasons, and so it appears there are reasons to weight reasons. With a regress looming, Schroeder knows that he will need a sophisticated account of weighting to tell a complete story, and so he makes several shifts in the ways we think about weighting and weight. Although his view is quite clearly more complex and nuanced than I can represent here, I will do my best to represent it fairly, even if leaving out some of the dialectic maneuvers.

The basic reorientation that Schroeder offers is based on two conceptual changes to common views about weight: first, reasons don’t have weights – sets of reasons do; and second, there is no ‘weight’ to be explained in the account, but instead there is a ‘weightier than’ relation. So the claim that ‘my reason to brew coffee is very weighty’ is sloppy in two regards: it is not the single reason which has a weight, but rather the set of reasons to do some thing; and there is no ‘great weight’ behind the activity of brewing, but rather we must be saying that my set of reasons to brew coffee is significantly weightier than my reasons to do something else (or perhaps even
anything else). And so we can say, for instance, that a set of three of my reasons to brew coffee is weightier than my single reason not to brew coffee, or \( \{S, T, B\} > \{L\} \).\(^{11}\)\(^{12}\)

Armed with these clarifications, Schroeder returns to the vindication of his ‘truism’ from earlier, which he formalizes as

**The Attractive Idea:** For \( R \) to be a weightier set of reasons to do \( A \) than \( S \) (\( R > S \)) is for it to be correct to place more weight on \( R \) than on \( S \) in deliberation about whether to do \( A \) \((2007a, 130)\).

Now as noted already, for it to be correct to place weight seems to imply that there is a reason to place weight. In fact, correctness seems to imply more than that – if it is correct to phi, then it seems there is *more reason* to phi than to do otherwise. So \( R \) is weightier than \( S \) if the set of reasons to place more weight on \( R \) than on \( S \) in deliberation is weightier than the set of reasons to do otherwise. However, not just any reasons will do, as has been observed by many in the recent ‘wrong kind of reasons’ discussion: If someone puts a gun to my head and tells me to place more weight on the reason to look at my shoelaces during every conversation than to look my interlocutor in the eye, this is the wrong kind of reason to do my deliberative weighting in a

---

\(^{11}\) Where each capital letter represents a reason, each of which can then be represented by an ordered triple. So \( \{S, T, B\} \), for instance, is the set of reasons represented by the triples \(<\text{Sleepy}, \text{Travis, Coffee}> \) (read: ‘his sleepiness is a reason for Travis to drink coffee’, and similarly for the other two); \(<\text{Tasty, Travis, Coffee}>\); \(<\text{Break, Travis, Coffee}>\). While reasons are *not*, in fact, ordered triples (but are facts that stand in the reason relation with an agent and an action), representing them as triples is helpful, Schroeder thinks, because this can help us to see which sets are ‘addable’ and comparable: namely, those which are invariant with reference to the agent and action \((2007a, 127)\). So \( S, T, B \) and \( L \) constitute a variety of addable sets, such as \( \{S\}, \{T\} \) and \( \{B\} \), which are addable into \( \{S, T, B\} \). This set is comparable with \( \{L\} \), which reads ‘that it will make it hard to fall asleep later is a reason for Travis not to brew coffee’.

\(^{12}\) This is a summary of large parts of Chapter 7 \((2007a, \text{esp. 123-139})\).
particular way. And so the weightier than relation must actually be analyzed in terms of the right kind of reasons to place more weight on one set of reasons than another.

At this point, there is an obvious threat of infinite regress. Since R is weightier than S just in case the set of the right kind of reasons to place more weight on R than on S in deliberation is weightier than the set of reasons to do otherwise, we now need to know about the weight of our reasons to place weight on our reasons to do some action. And it will be correct to place weight on the reasons to place weight on the reasons to do the thing just in case the set of reasons to place weight in that way is weightier than the set of reasons to do it some other way. And so on.

In response to the threat of regress, Schroeder formulates the disjunctive ‘Recursive Analysis’. On this account, the weightier than relation can be analyzed as follows:

**Weight Base:** One way for set of reasons A to be weightier than set of reasons B is for set B to be empty, but A non-empty.

**Weight Recursion:** The other way for set of reasons A to be weightier than set of reasons B is for the set of all the (right kind of) reasons to place more weight on A to be weightier than the set of all the (right kind of) reasons to place more weight on B (2007a, 138).

The idea here is very clever, and it solves almost all of Schroeder’s problems. He is able to capture the intuition that weight is normative, but without it resulting in the infinite regress that was looming. How? Well, because it doesn’t avoid the *regress*, just the infinite part. The

---

13 For just a small sample of the extensive literature on the Wrong Kind of Reasons Problem, see D’Arms and Jacobson (2000), Olson (2004), Rabinowicz and Rönnow-Rasmussen (2004; 2006), and Schroeder (2007a; 2010; 2012).
insight here is that in the case of placing weight in deliberation, our reasons will eventually run out. And when they do, the case will be the one described in Weight Base – there will be reasons to place more weight on one set of reasons, and no reasons to place weight in another way. So every case of deliberation must ‘bottom out’ in a substitution instance of Weight Base, which then explains at each step of the recursive normative chain why the set of reasons to place weight on one set of reasons is weightier than another.

There is no question that Schroeder’s Recursive Account is an interesting attempt to get the named desiderata out of a version of HTR. However, I will argue that the view ultimately fails to be a plausible candidate theory of weighting. Buried deep within the technical maneuvers of The Recursive Account is a problem that renders the view either circular or ad hoc.

3.5. The Failure of The Recursive Account

Given the importance of Schroeder’s arguments to the literature on reasons, his discussion of weighting has already received significant attention and criticism. Some of these criticisms provide good reasons to be skeptical of The Recursive Account, but since the points have been made elsewhere, I will not focus on them here. Instead, I here want to challenge The Recursive Account in a way that I do not think it has been challenged, and which seems to me quite powerful. In short, because The Recursive Account attempts to make sense of the ‘truism’

---

14 See, for instance, the discussion between Evers (2009), Schroeder (2007a; 2010) and Enoch (2011) on Schroeder’s solution to the Wrong Kind of Reasons Problem. Although Schroeder makes some interesting gestures towards solving this problem, he does not provide a complete solution, and yet his position requires that there is one. So one concern for his view is that it requires a solution to a problem that is very difficult, and that one might even be skeptical of solving.
that weight is normative, it fights constantly to avoid regressive and circular analyses. If we are careful to understand the view correctly, then, we will see that even the part of the analysis that is supposed to put an end to the main regress is, itself, in need of a recursive analysis, and stopping this new regress requires either circular reasoning or answer by *fiat*, both of which I take it are unsatisfying.

The key to understanding The Recursive Account – and to what I take to be its structural implausibility – is the obvious fact that it is a *disjunctive* account. There is one analysis of the weightier than relation – Weight Recursion – that follows directly from the intuition motivating the view. However, there is a serious (devastating) problem with Weight Recursion as a solo analysis – namely, it entails an infinite regress. So Weight Base is posited to stop the regress. And while Schroeder admits that The Recursive Account is unambiguously disjunctive, and that he, himself, takes an account’s being disjunctive as a weakness, he states that he is less worried about this than he might be, since the two halves of the account are unified by their making sense of The Attractive Idea with which we began (2007a, 139-140).

Before analyzing this defense, one might also defend the disjunctive aspect of The Recursive Account by claiming that Weight Base is so perfectly intuitive that it is obviously true regardless of what else is the case. And in a sense, I agree. If sets can be weightier or less weighty than one another, I suppose I would expect it to be the case that the empty set is less weighty than non-empty sets. However, here I am sympathetic with Enoch’s criticism that our intuitions seem to be getting unreliable at this point. Perhaps I am imagining a scale, on which I am measuring something against nothing; but in my imagination, the *something* is heavier

*because it has weight*. As Enoch puts the point:
For recall that Schroeder has rejected the thought that weights are magnitudes that reasons come with. Rather, for one set of reasons to be weightier than another is (roughly) for it to be correct to place more weight on it in deliberation. But Weight Base looks so natural only, I submit, because we implicitly still hold on to the magnitude view. If weight is a magnitude, then clearly something is always weightier than nothing. But once we’ve deserted this view, as Schroeder does, nothing about Weight Base seems obvious any more (2011, 441).

On Enoch’s view, this is the ultimate problem with Weight Base, and I admit that it is a problem with Weight Base. However, stopping the conversation here is unfair to The Recursive

15 One might object that the magnitude view of weight is not needed in order to vindicate our intuitions about Weight Base. Instead, one might argue, all that is required in order to see that Weight Base is true is that it turn out to be the case that all reasons have weight (and so, a set with any reason in it is weightier than the empty set). But one could derive this claim from Schroeder’s view. While the ‘magnitude view’, which Schroeder rejects, holds that reasons have weight, and the weightier than relation is determined by the magnitude of these weights, an alternative view would switch the order of explanation. Perhaps the weightier than relation is explanatorily primary, as Schroeder claims, but it is also the case that all reasons have weight. How? Well, one could take equivalence classes to determine how much weight a set of reasons has. For instance: consider the equivalence class of sets that are weightier than the empty set, but not weightier than any non-empty set. We could say of each of these sets that they have the minimal amount of weight; if we were interested in assigning numerical values to work with, each of these sets could be said to have weight ‘1’. All non-empty sets not in this equivalent class thus has a weight > 1.

There are at least two problems with using such a view to explain Weight Base. The first, and most obvious, is that Weight Base is part of the explanation of the weightier than relation, but the move under investigation uses the weightier than relation to derive an explanation of Weight Base. But this sounds like a circular explanation. It can’t be the case that the relation being explained by Weight Base is precisely what explains it. The second problem with this explanation of Weight Base is that it undermines the unity of the Recursive Account. Schroeder claims that the disjunctive nature of the account does not bother him because the two disjuncts are unified in their explanation of The Attractive Idea. However, if Weight Base is true because all reasons have weight in a derivative sense, then the disjunctive worry comes back. Such an account would employ two different explanations for two different components of the
Account, as Schroeder knows perfectly well that ‘the magnitude view’ of weighting cannot be what makes Weight Base plausible. So on the one hand, I want to demand more charity for The Recursive Account, and to require that we have a very full, nuanced understanding of the view before evaluating its success, and Enoch hasn’t accomplished this. However, my demand here is a bit underhanded, as I will argue that a full understanding of The Recursive Account makes the view less, not more, plausible.

Weight Base, then, is stranger than it looks. It claims that a non-empty set of reasons is weightier than the empty set, but that the weightier set is not weightier in virtue of having any weight. So why should we believe Weight Base? Schroeder thinks that there is in fact a good reason to believe it, and that is because Weight Base is the claim which, when combined with Weight Recursion, yield The Attractive Idea. Now The Attractive Idea, recall, is just the beginning intuition that for set of reasons $R$ to be weightier than set of reasons $S$ is for it to be correct to place more weight on $R$ than on $S$ in deliberation. And so Schroeder’s attempt to defuse the disjunction worry is by pointing to the way in which we can see the two disjuncts as working together to yield this intuitive result.

The problem, however, is that Weight Base and Weight Recursion don’t get us The Attractive Idea. And this is precisely the deeper worry that I think is ultimately pointed to by Enoch’s objection above. The Weight Recursion half of the disjunct clearly contributes to explaining The Attractive Idea, as it entails The Attractive Idea, given (as Schroeder believes) that it is correct to do something just in case the set of reasons to do that thing are weightier than theory. Thanks to Steve Kuhn both for raising this possibility, and for helpful discussion of its viability.
any other. But does Weight Base get us The Attractive Idea? No. The Attractive Idea is about the correctness of placing weight in a particular way; but Weight Base involves no placing of weights.

In order for Weight Base to be anything more than *ad hoc*, something else must be true, and Schroeder knows this. He tells us that we get the desired conclusion in Weight Base – that it is correct to place more weight on A (the non-empty set) than on B (the empty set) – “if we make an assumption that I think is eminently plausible – that there is a general reason to place more weight on reasons than on non-reasons in deliberating about what to do, and that there is no such reason the other way around” (2007a, 140). So in order to make the disjunctive account predict the motivating intuition, and to thereby keep it from being objectionably *ad hoc*, we need to assume an agent-neutral reason about how to deliberate.17

At this point, I think we have an understanding of the total structure of The Recursive Account and particularly Weight Base. Now let’s return to Enoch’s worry. As formulated, the

---

16 It has been suggested to me by various commenters that the locution ‘to place weight on non-reasons in deliberation’ is at least odd, and perhaps incoherent. If one thinks, for instance, that deliberation just is the activity of placing weight on reasons, or that reasons just are the kinds of things that one can weight in deliberation, it is totally unclear what this phrase can mean. I agree that it’s odd; however, by using it here, I am not endorsing this way of speaking. I am simply engaging with Schroeder on his own terms. In my positive theorizing in what remains, it will not be necessary that one can place weight on anything other than reasons (and in fact, I will, below, take issue with the very idea of ‘placing weight’ – what sounds like a *practical* activity – as what one does in deliberation).

17 One might already be worried at this point that, since this is a Humean account, this reason to place weight must be grounded in desires, and it must be a reason for everyone. While I do, in fact, find it implausible that a Humean account can generate a universal reason to place weights in this way, such a discussion is a larger one than I can have here. My suspicion at this point is due to the implausibility of generating agent-neutral reasons from desires generally. For the details of Schroeder’s view, see (2007a, ch. 6).
concern is that Weight Base seems true only because we implicitly believe the ‘magnitude’ view of weighting, but Schroeder has abandoned that view. However, Schroeder is not caught up quite that easily, as he sees this problem and has a built-in response in the form of the general reason to place more weight on reasons than on non-reasons. But now I think that Schroeder has given away the game.

The entire job of Weight Base was to stop the regress, and it no longer does that. When we look at the disjunction that makes up The Recursive Account, we see that Weight Recursion faces an infinite regress, with reasons all the way down. But then Weight Base is there to stop the regress, which it does by explaining the weight of reasons without reference to reasons. Sure, says Schroeder, one way that set of reasons A is weightier than set of reasons B is for there to be reasons to place more weight on A than on B, and this is troubling because there are still reasons with weight unaccounted for in the explanans. But no worries, because when we finally hit a weight base case, A can be weightier than B just by A being non-empty and B being empty. Nothing normative, no reasons in the explanans, and so no more weighting that needs to be explained by more reasons. But by admitting that Weight Base isn’t actually the bottom – that there is a more fundamental explanation – he reintroduces reasons into the explanatory story, and so the regress begins anew.

More carefully: Weight Base is true, not because non-empty sets have weight, but because there is a reason to place more weight on reasons than on non-reasons (for ease of discussion, let’s call this The Bottom Reason). But now it looks like The Recursive Account faces a dilemma. Either The Bottom Reason has weight or it does not. The idea that any reason can be weightless is at least odd, and perhaps incoherent. After all, a reason is a normative unit,
and weight is how the normativity of that unit is measured. So a weightless reason would be simultaneously normative and without normative force. It seems clear that we should therefore reject the second horn of the dilemma. So The Bottom Reason must have weight. *But that means that there must be reasons to place more weight on The Bottom Reason than to do anything else.* And so our regress continues.

It looks to me, at this point, like we should expect that the explanatory regress is unstoppable, and that this would be the best reason to reject The Recursive Account. However, the dialectic has not quite run its course. One very good piece of evidence that Schroeder saw this criticism in advance is that he not only posits The Bottom Reason as the explanation for what makes Weight Base true; he also posits that there are no reasons ‘the other way around’, which is to say that there are no reasons to place more weight on non-reasons than on reasons. Why would he say this? Well, because it makes The Bottom Reason an instance of Weight Base. The set of reasons to place more weight on non-empty sets of reasons than on empty sets of reasons is, itself, a non-empty set of reasons up against an empty set. This makes it look to me as if Schroeder thinks he can avoid my new regress challenge by showing that The Bottom Reason is a Weight Base case, and so the set of reasons made up of The Bottom Reason is weightier than empty sets of reasons.

Although this seems to be the only explanation for why Schroeder insisted both on the existence of The Bottom Reason and on the non-existence of competing reasons, such a response is unconvincing. Because The Recursive Account has abandoned the magnitude view of weight, we are entitled to demand an explanation of what makes Weight Base true. Now either The Bottom Reason is supposed to explain Weight Base or it isn’t. If it is, then it cannot employ
Weight Base in the explanation. It is clearly circular to say that Weight Base is true because there is a reason to place weight on non-empty sets, which itself is a Weight Base case, and so is weightier than non-reasons because Weight Base is true. But if The Bottom Reason isn’t really supposed to explain what makes Weight Base true, then we are left without any reason to believe Weight Base. It is not plausible for the intuitive reasons, as those are based on a magnitude theory of weight. But it is also not plausible for reasons internal to The Recursive Account, as it continues to sit, unexplained, at the bottom of the explanatory story.

At this point, I believe I have shown that the view of weighting put forward by Schroeder in *Slaves of the Passions* ought to be rejected. It avoids an infinite regress only through either circular reasoning, or by employing an *ad hoc* stipulation. However, one might think that the Recursive Account is sufficiently promising that we ought to look to see whether it might be saved by rejecting Schroeder’s actual position, and so I will now evaluate one maneuver for avoiding the challenge raised here.

3.6 Can The Recursive Account be Saved?

According to my criticism, Schroeder’s argument is doomed when he introduces The Bottom Reason. By claiming that there is a reason to place weight on non-empty sets, and no reason the other way around, Schroeder requires either (a) that there is reason to place weight on The Bottom Reason, and so the regress begins anew, (b) that The Bottom Reason has weight by virtue of being an instance of Weight Base, which would make the view circular, or (c) that we ignore The Bottom Reason, and make Weight Base an *ad hoc* stipulation. Our question, then, is whether The Recursive Account has some other way of stopping the explanatory regress without
introducing The Bottom Reason. In what follows, I will investigate one final defensive maneuver that a defendant of the Recursive Account might explore.\footnote{Thanks to Mark Murphy for suggesting the following line of reasoning.}

Perhaps non-empty sets are weightier than empty sets, because there are, in fact, reasons to place weight in that way, but that these reasons come from \textit{within the non-empty set}. Perhaps any reason to do anything is \textit{also} a reason to place more weight on a non-empty set than on an empty set. On such an account, when it came to an instance of Weight Base, we would have two sets, \{X\}\footnote{Although I here use a particular case of a singleton set being compared to the empty set, this of course is not necessary; the non-empty set could have any (non-zero) number of reasons. I here use the particular example merely for ease of explanation, since the account I am sketching here requires that we discuss particular reasons within the non-empty set.} and \{ \}, and X would also be a reason to place more weight on \{X\} than on \{ \}. It would thus be the case that there would be some reason to place more weight on \{X\} than on \{ \}, and no reason the other way around, and so it would be true that \{X\} \textgreater \{ \}. Call this the Circular Defense, as the reason to place more weight on \{X\} than on \{ \} comes from within \{X\}.\footnote{Although the justification is clearly circular (hence the name), it doesn’t seem to me to be viciously circular, and so the ‘circularity’ will not constitute any part of my worry for the view.}

I do not know whether Schroeder would even want to go this route. Perhaps he considered such a view and rejected it, either for some of the reasons I will here consider, or for others that I haven’t discerned. In any event, I will close this section by raising some worries that I have about such a defensive move. Perhaps, at the end of my defense of proportionalism, someone will believe that reviving The Recursive Account along these lines really is preferable to simply adopting proportionalism. I find such a prospect unlikely, but if one were so moved to argue for The Circular Defense, she would need to address the following issues.
Most importantly, we need to get clear on the details of how The Circular Defense works, and what else one would need to endorse in order to embrace it. The central idea of the defense is that the reason to place more weight on non-empty sets than on the empty set comes from within the non-empty set under investigation, so that there is no need to posit Schroeder’s Bottom Reason. This is supposed to be helpful in stopping the new regress, because it draws a tight, justificatory circle from within any non-empty set, such that any reason to do anything is also a reason to weight that reason in deliberation, and so there is a reason to place more weight on any non-empty set of reasons than on the empty set. It thus clearly seems right that the non-empty set is weightier than the empty set, by virtue of having some (any) reason to place more weight on it than on the empty set, while there is no such reason the other way around.

A full investigation into the defensibility of this view – which clearly is not Schroeder’s – would take more space than I have here. But I do want to raise two issues that I take to cast serious doubt on the defensibility of such a maneuver. The first is a formal one, as I am not quite sure that we can generate a reason to place more weight on non-empty sets that would be inside the relevant non-empty sets. The last is a more substantive worry about what the reason to place weights in this way would be. I will address each in turn.

In my summary way of putting The Circular Defense, I claimed that perhaps the reasons that make up a non-empty set are also reasons to weight non-empty sets in one’s deliberation. But what might be meant by this claim that some reason ‘is also’ another reason? To the extent that it was an intuitive thing to say to begin with, I take it that we ought to read the claim as follows: the fact that stands in the reason relation to constitute X, also stands in the reason relation to constitute X’ (where X’ is a reason to place more weight on non-empty sets than on
the empty set in deliberation). 21 The locution that some reason ‘is also’ another reason is sensible because the fact that constitutes the reason is capable of standing in the reason relation to multiple agents and actions, and so some fact can easily constitute both X and X’.

For example: I have a reason to continue writing, constituted by the fact that writing is necessary to finishing my dissertation; the relevant background desire, then, is my desire to finish my dissertation, or my desire to graduate. However, the fact that writing is necessary to finishing my dissertation is also a reason to place more weight on my set of reasons to keep writing than on the empty set, as doing so (it might seem) promotes my actual writing, and so my desire to finish my dissertation or to graduate. 22 And so if (what is likely impossible), I had only this one reason to continue writing (X), and no reason not to write, then the explanation of why \{X\} > \{\} is that the fact constituting X also constitutes another reason (X’), and X’ is what places more weight on \{X\} than on \{\}.

In drawing out how the example works, the first structural issue may now seem obvious: I started out considering a non-empty set \{X\} (a singleton set, constituted by one reason, namely X), but the circular defense only works if the reason to weight \{X\} comes from within \{X\}; so it

\[\text{\footnotesize 21 Although slightly cumbersome, I use X and X’ to represent the two reasons quite intentionally. On the one hand, a reason is simply the consideration that stands in the reason relation to some act and agent; as a result, we should expect the ‘two reasons’ being discussed here to actually be the same reason – that is, one reason to do two different things. However, we should also expect it to show up twice, because there are two normative facts involved: a reason to act, and a reason to weight sets of reasons. So the reasons are both ‘X’, as they are the same reason (consideration), but there are two reasons to do two different things, and so the second is marked as prime.}
\[\text{\footnotesize 22 Whether this is a good model for generally explaining the reason to place weight on non-empty sets will be explored further below. For now, I simply need an example, and this seems like a fairly intuitive explanation, it seems to me.} \]
appears that \{X\} must actually be \{X, X’\}. On the one hand, this doesn’t look problematic, as X and X’ are the same reason – that writing is necessary for finishing my dissertation – for two different actions. But if they’re actually the same reason, then \{X, X’\} is actually just \{X\}, so we’re back where we started. However, as I noted in f. 21 above: although X and X’ are the same reason, they are the same reason to do different things. Thus, there are two normative facts there, and so they shouldn’t be reducible to a singleton set.

The confusion concerning whether X and X’ are one reason or two is due, I think, to the idea that X’ could ever be in the original set to begin with. In short: X and X’ aren’t different because they’re different reasons; they’re different because they’re reasons to do different things. As a result, they will show up in different sets from one another. To see this, let’s turn towards Schroeder’s discussion of the ‘addability’ of sets. \{X\} and \{Y\} are addable sets, in Schroeder’s sense, if the agent and action standing in the reason relation with X and with Y are invariant (2007a, 127). This is just to say: X and Y constitute addable sets if they are reasons for the same agent to do the same thing. However, \{X, X’\} is not addable with any set; in fact, \{X\} and \{X’\} are not addable. Why? Because X and X’ are not invariant with respect to action. Going back to our example, X is a reason for me to write my dissertation, and X’ is a reason to place more weight on my set of reasons to write than on the empty set. But reasons to write dissertations and reasons to weight reasons are not the kinds of things that can be agglomerated. This seems to suggest that we do not, in fact, want X’ to be inside the set being weighted against the empty set in Weight Base; but of course, putting X’ outside of that set takes us back to Schroeder’s solution, according to which we have a general reason to place weight on non-empty sets.
I don’t know if this issue could be successfully addressed through fancy, formal footwork. Perhaps one could attempt to make plausible a different sense of ‘addable’ and ‘comparable’ sets, according to which the relevant ‘prime reason’ (X’ in my case, Y’ in the case involving the reason Y, etc) accompanies every reason to do anything. But the argument sketch above makes plausible, I think, the following dilemma: a reason to weight non-empty sets must either be in the non-empty set under consideration, or not. If not, then the view is basically Schroeder’s, and it faces the criticism that I have leveled against his version of The Recursive Account. If the reason to weight non-empty sets is in the non-empty set under consideration, however, then it makes it difficult to see how our sets of reasons for action are addable and comparable, as all sets will be constituted by reasons to do at least two, different actions: that is, all sets will be constituted by reasons to phi, and by reasons to weight non-empty sets in deliberation.

Suppose that someone is able to convince us, however, that this last result is not problematic. Perhaps a sufficiently nuanced theory of addition and comparison among sets will reveal such a view to be defensible. The Circular Defense is now faced with the important question: what is W, such that all facts that constitute any other reason also constitute it?

In the dissertation example above, I alluded to the only explanation that I could imagine that would make The Circular Defense true, and it runs as follows: for any reason (A) to phi, the fact that constitutes A also constitutes a reason (W) to place more weight on {A} than on { } in deliberation, because placing more weight on {A} than on { } in deliberation promotes actually phi-ing. So the fact that writing is necessary to finishing my dissertation is a reason to write (X). However, the fact that writing is necessary to finishing my dissertation also constitutes a reason
(X’) to place more weight on \( \{X\} \) than on \( \{\} \) in deliberation, as doing so promotes actually writing. So we have one reason to do two things, represented by the triples, \(<\text{Writing is Necessary, Travis, Write}>\), and \(<\text{Writing is Necessary, Travis, Weight}>\). So long as we can solve the formal addability problem raised above, then, this seems like a plausible explanation of X’. It does, pre-reflectively, seem like placing more weight on my set of reasons to write than on the empty set promotes my actual writing. However, the last worry I want to mention before closing is that, plausible though it may sound, there is a class of cases in which placing more weight on some set of reasons to act does not promote that action. These ‘self-defeat’ cases reveal that The Circular Defense, even if it could be made to work formally, does not generate as wide-spread a reason to place more weight on non-empty sets as one might hope, and so The Recursive Account is left without a complete explanation of Weight Base.

Consider Schroeder’s case of Nate, who loves successfully thrown surprise parties in his honor, but who hates all other kinds of parties (including failed surprise parties in his honor). Schroeder tells us about Nate that, given his quirky desires, “the fact that there is a surprise party waiting for him now at home is a reason for him to go home. But it isn’t a reason that Nate could know about or act on” (2007a, 33). The case of Nate is interesting for many reasons, but for present purposes, it is most important as a counter-example to the claim that placing more weight on a set of reasons in one’s deliberation promotes the action favored by those reasons. In Nate’s case, the fact that there is a surprise party waiting at home is a reason, \( Y \), to go home, and so Weight Base tells us that \( \{Y\} > \{\} \), such that, if Nate had no other reasons to act relative to going home, he would have weightier reason to go home than to not go home. The Circular Defense then holds that the fact that there is a surprise party waiting at home actually constitutes
another reason – a reason to place more weight on \{Y\} than on \{\} in deliberation – because doing so promotes actually going home. But in the case of Nate, this isn’t true. Weighting a set of reasons in deliberation requires taking oneself to have those reasons, but Nate cannot take himself to have the reason Y. If he knew that there was a surprise party waiting for him at home, the reason to go home would evaporate; in fact, since he hates failed surprise parties, he would then have a reason not to go home. So weighting \{Y\} in his deliberation, which requires taking himself to have Y, would in fact not promote Nate’s going home – it would undermine, or defeat the goal of actually acting in accordance with Y. So there is at least one case in which the intuitive case for believing that all reasons, A, imply reason W, is undermined. But of course the structural features of Nate’s case aren’t unique. Any reason that exists only when the agent for whom it’s a reason doesn’t know about it will generate a case analogous to Nate’s. And in all such cases, if there are no other relevant reasons (making it an instance of Weight Base), The Circular Defense will fail to explain why non-empty sets are weightier than empty sets.

This concludes my very abstract criticism of a very abstract theory. It may seem as if I have driven into the tall weeds, trying to score merely technical points against a grand theory. But the discussion – although perhaps technical – is not merely technical. If I am right, then Schroeder’s version of The Recursive Account is either ultimately circular, or there is no reason to believe the Weight Base half of it. But if there is no reason to believe the claim that was posited specifically to stop a regress problem, then it becomes very difficult to see how the view is not (contrary to Schroeder’s appeals) ad hoc. Further, while The Circular Defense provides one option for reviving The Recursive Account, I have been unable to determine both how the formalism of the defense can be made to work, as well as what could explain the substantive
claim that all reason-constituting facts constitute an additional reason to place more weight on non-empty sets than on the empty set in deliberation. I thus conclude that a Humean ought to reject The Recursive Account as a theory of weighting.

3.7 Defending Proportionalism

In favor of proportionalism, my arguments so far have claimed both: (a) that it is clearly the most natural theory of weighting; and now (b), that The Recursive Account – both as Schroeder articulates it, and utilizing The Circular Defense – should be rejected. However, this is not yet a complete victory. According to Schroeder’s argument rehearsed in 3.3, the Humean is pressured towards The Recursive Account because proportionalism does not explain the way in which the weight of reasons is a normative matter, and so it fails to vindicate the ‘truism’ that it is always correct to place more weight in deliberation on a weightier reason. I have not successfully argued for proportionalism, then, until I have defended it against Schroeder’s accusation. Doing so is the goal of the current section.

On the one hand, Schroeder’s claim that the ‘weight of reasons is a normative matter’ is perfectly intuitive. After all, reasons are the basic unit of normativity, and weight is the measurement of that normativity, as one has more or less reason to act as a result of one’s reason to so act having more or less weight. So weight is clearly normatively relevant. Further, one of the ways in which it is obviously normatively relevant is that something like Schroeder’s ‘truism’ does, in fact, seem true; that is, weight seems normatively relevant because it seems correct to place more weight on a weightier reason in one’s deliberation. So far, it sounds like I am in agreement with Schroeder. My modest claims here, however, are not what Schroeder infers from the claim about the normativity of weight. Instead, he goes on to say that “all
normative properties and relations are analyzable in terms of reasons, for that is what makes them normative. If that is right, then claims about the weight of reasons need to themselves be analyzed in terms of reasons” (2007a, 100). And this is where I think it is perfectly intuitive to resist. Because Schroeder subscribes to ‘Reasons Basicness’, he holds that “what it is to be normative, is to be analyzed in terms of reasons” (2007a, 81), and so if weight is normative, then it is, in fact, to be analyzed in terms of reasons. On such a view, something like Schroeder’s truism falls out as an explanation of the weightier than relation.

Notice that the claim that ‘the weight of reasons is a normative matter’ is distinct from the claim that weight is to be analyzed in terms of reasons; the latter requires the former plus Reasons Basicness. Thus, one clear way to resist the implication that the weight of reasons is to be analyzed in terms of reasons is simply to reject Reasons Basicness. Although the view is popular among contemporary theorists, it is not obvious, and many will be comfortable resisting it. Proportionalism might be the correct account of weight, and it might be that weight is a normative matter, because being normative may not imply being analyzable in terms of reasons.

Another option is to focus on something I wrote in my timid agreement with Schroeder: that the weight of reasons is clearly normatively relevant. What’s interesting about this language, which seems very natural to me, is that Schroeder himself discusses the distinction between facts that are ‘normatively relevant’ and those that are ‘normatively contentful’. Only

23 Schroeder notes about Reasons Basicness the following: “Any view that looks attractive to Jonathan Dancy, Jean Hampton, Michael Smith, Derek Parfit, T. M. Scanlon, and Joseph Raz has to be one with some kind of very broad appeal” (2007a, 81).
the latter, he says, constitute facts that are ‘normative’ in the sense of being analyzable in terms of reasons. The example of a normatively relevant fact given by Schroeder is the fact that pointing a loaded gun at my head and pulling the trigger will surely result in my death. Such a fact is normatively relevant, says Schroeder, because it implies that one has an exceptionally good reason not to do those things. Normatively relevant facts thus imply, entail, or have as a consequence various other facts about reasons; but these facts are not, themselves normative. Facts about guns and death, Schroeder writes, should not be thought of as normative (2007a, 80). Another option for the proportionalist, then, is to note that weight is, as I wrote, clearly normatively relevant, as it plausibly entails normative facts, but that this does not imply that weight is normative.

Since I do not want to assume the falsity of Reasons Basicness (although I am also not assuming it is true), I will pursue the latter strategy for explaining the sense in which weight ‘seems normative’. But whether one rejects Reasons Basicness, or claims that the weight of reasons is normatively relevant, rather than normative, Schroeder claims that if one accepts proportionalism as a theory of weight, she incurs the burden to explain a truism that simply falls out of The Recursive Account. That is, even if one is not forced by the acknowledgement that weight is normatively relevant to Schreoder’s view, he claims that proportionalism “is ill-suited to explain the truism that it is always correct to place more weight on a reason that is weightier. If Proportionalism is the right analysis of the weight of reasons, then this truism is completely mysterious” (2007a, 101; emphasis in original). My response, however, is that once we acknowledge that a fact can be normatively relevant without being normative, I see no reason to
accept Schroeder’s claim that proportionalists must find the truism ‘completely mysterious’. Indeed, I think the truism is true, and for fairly straightforward reasons.

The claim that it is always correct to place more weight on a weightier reason can be read in at least two ways: it can be read as saying something about what practical reasons there are, or about what epistemic reasons there are. I take it that Schroeder opts for the former strategy, which is part of why his analysis of the weightier than relation invites a regress: when I have a reason to write my dissertation, I also have a reason to do something else – namely, to place on the reason to write in my deliberation. Although it is unclear to me exactly what this activity of ‘placing weight’ consists in, it does sound like an activity – like I something that I do. I do not, however, think of deliberation as an activity in which I do things to my reasons. That is: I do not think of deliberation as a practical activity. Rather, I think of deliberation as an epistemic activity, in which I discover things about my reasons, take myself to have reasons of some weight or another, or believe about reasons that they have some weight. On this picture – what we can call the Epistemic Model – my interaction with my reasons has a different direction of fit from that which happens on Schroeder’s Practical Model.

How does the move from the Practical Model to the Epistemic Model help? On the one hand, I will have to admit at the end that it pushes the conversation back a step, into deep, important questions that I cannot answer here. On the other hand, the move to epistemic language allows me to reword Schroeder’s truism in a way that both seems more accurate, and completely non-mysterious. On an epistemic reading of deliberation, what seems clearly true about the weight of reasons is that it is always correct to believe about a weightier reason that it is weightier. And why should that be true? For the very simple reason that true belief is correct
belief, so when it is true that X is a weightier reason than Y, it is correct to believe that X is weightier than Y. This explanation, I take it, is incredibly intuitive and non-mysterious, and is perfectly compatible with proportionalism.

I did say that the move to an Epistemic Model of deliberation would push back some deep, important questions, however, and I will just briefly mention what I mean by that. Although I rephrased Schroeder’s truism in a way that allowed me to explain ‘correctness’ within the epistemic domain, my opponent may challenge that I have merely exchanged a puzzle about practical reasons for one about epistemic reasons. After all, if it is correct to believe about a weightier reason that it is weightier, doesn’t this imply that there is reason to believe about a weightier reason that it is weightier? But if so, what weight does this reason to believe have? And won’t it be true about that reason that, if it is weightier than some other reason, that one will have reason to believe that? The worry here is that I have invited a regress similar to the one that I used against The Recursive Account.

In response, it is important to note first of all that the proliferation of epistemic reasons just cited does not constitute an argument that proportionalism, as I have described it, faces a regress problem. Crucially, explaining how the truism is (or at least appears) true is not the same as appealing to it as explanatory, and so proportionalism has not been shown to face an infinite explanatory regress. Rather, something else is going on. The worry amounts to something like the claim that, if epistemic correctness implies a reason to believe, then it appears that there must be an infinite number of reasons to believe (since a reason has weight, which will imply a claim about correctness, which will imply a claim about weight, etc). This is an interesting observation, and although it is a new challenge to the proportionalist Humean, it does appear
troubling. However, at this point in discussion, what I want to point out is that my interlocutor has pushed me further than seems fair; that is, whether or not this criticism is harmful to my position depends on the details of what may turn out to be an entirely different theory – one that explains epistemic reasons.

Recall that the Humean position being defended in this chapter is a theory of practical reasons; indeed, I (unlike Schroeder and other Humeans) have allowed an even more modest reading of HCA, on which at least some reasons seem explainable in the way that Humeanism claims. Thus, my defense of a proportionalist Humeanism here claims only that (at least some) reasons for action are explainable by background desires, and that the weight of these reasons is proportional to the strength of those desires. I have certainly not committed myself to any claims about how epistemic reasons work.

So perhaps the challenge is that I need a view of how epistemic reasons work, in order to completely defend my proportionalism.\(^{24}\) However, such a request seems too demanding. Perhaps if proportionalism obviously implied a view of epistemic reasons that was clearly indefensible, then it would seem that I had inherited an explanatory burden regarding them; but the few comments above seem to fall short of justifying so strong a claim. Instead, it seems to justify something like the claim that a proportionalist ought to think more about what it wants to say concerning epistemic reasons, given this interesting consequence. And that seems true. Not

\(^{24}\) This would not be the only criticism to claim to tie the fate of HTR (or some other, similarly reductionist account of practical reasons) to the details of an account of epistemic reasons. Several philosophers have recently given ‘companions in guilt’ arguments against any antirealist position, invoking reasons for belief as understandable only on a realist metanormative model. See, most notably, Cuneo (2007), but also Jackson (1999), Putnam (1990), Scanlon (1998), and Shafer-Landau (2003).
only do I hope to say more about the issue in other work, but I have a suggestive idea, with which I will close.

The worry about infinite proliferation of epistemic reasons gets moving because it makes a seemingly innocuous assumption: that if it is correct to believe something, that one has reason to believe that thing. And the same assumption was made in the practical domain (with Schroeder, in fact, endorsing the stronger position that for something to be correct to do is for there to be most reason to do it). However innocuous such an assumption may be, I am unsure that we should endorse it. While ‘correct’ seems in the practical case, especially to one who endorses Reasons Basicness, to mean something like ‘favored by reasons’, the different direction of fit in the epistemic case makes it less clear that ‘correct’ has such a meaning. It may be correct to believe something simply because it’s true, and belief seems to be an activity that ‘aims at the truth’. This sounds intuitive. And it may also be the case not all truth is reason-giving.\textsuperscript{25} That is, it may be that the correct theory of epistemic reasons allows for correct beliefs that there is no reason to hold (and, perhaps more interestingly, incorrect belief that there is reason to hold). That this seems plausible suggests that the move from correctness to reasons to believe was too fast.

Of course, completely spelling out how exactly it is that all of these things are true about epistemic normativity would require its own project, and so I leave off for now. For my

\textsuperscript{25} One view that would make such a claim is a global desire-based view, on which all reasons – practical and epistemic alike – are grounded in the desires of agents. Schroeder, in (2007a) actually admits that he thinks something like this is probably true, but does not argue for it. And Street, although not a \textit{desire}-based theorist, agrees that a theory on which reasons are dependent on the agent’s psychology is likely the correct account for all reasons – practical and epistemic alike (2011).
purposes, what is important is that proportionalism need not deny Schroeder’s ‘truism’ – only that it provides an analysis of the weightier than relation. In fact, proportionalism can (and should, I think) accept the truism, understood according to the Epistemic Model – that it is always correct to believe about a weightier reason that it is weightier. And while the introduction of epistemic reasons has the potential to raise new challenges, it does not seem fair to require of proportionalism that it have a complete theory of epistemic reasons worked out. Given that my defense of proportionalism hasn’t entailed anything obviously implausible about the epistemic domain, it is sufficient to keep in mind that adopting a proportionalist Humean view of practical reasons may have ramifications for theorizing about epistemic reasons as well.

3.8 The Evaluation

So why be a proportionalist? I claimed at the beginning that the answer was rather straightforward, and so it is: the view is prima facie plausible for a Humean, and there are no good reasons to resist that plausibility. We can now see in more detail how this argument works: first, the very intuitions that motivate HTR in the first place seem to motivate proportionalism. The question, then, is why one should reject it. Schroeder claims that it fails to respect that fact that weight is normative, while his own Recursive Account entails the normativity of weight. However, I have argued that Schroeder’s view faces a devastating problem, as it either entails an infinite regress or is objectionably ad hoc, and that the only strategy I can see for salvaging the account also fails. This is not bad news for the Humean, however, because proportionalism also can explain the seeming normativity of weight, and accept Schroeder’s truism that it’s always correct to place more weight on a weightier reason; it’s just not the case that proportionalism
makes the truism *trivially* true. Thus, proportionalism is the natural view, and both Schroeder’s criticism of it, as well as his argument for the candidate replacement view, fail.

There is one other consideration that an opponent of proportionalism may want to raise, which is the belief by many that proportionalism makes HTR more difficult to defend. As mentioned in the beginning, Schroeder agrees with his opponents that if HTR is forced to accept proportionalism, then the resulting theory would be indefensible; in Schroeder’s language, the Humean would be unable to fully answer the charge that HTR implies that there are both too few and too many reasons (2007a, 101-102).

But this sort of worry is insufficient as a reason for adopting one theory of weighting over another. There must be something to recommend the view *from the perspective of HTR*; otherwise our choice of theory seems like dialectical cheating. As Enoch nicely puts the point:

If Proportionalism is a very natural thesis for a Humean to accept, and if it is not sensible (because there are Humeanism-independent reasons to reject it), then this does not count in favor of a Humean view that nevertheless rejects Proportionalism. Rather, it counts against Humeanism. When a challenge has the form of (roughly speaking) a reductio ad absurdum of your view, you cannot just retreat to being sensible, reject the absurdum (it is, after all, quite absurd!), and then fix in your theory whatever needs fixing to avoid entailing the absurdum. Were this enough, it would have been the perfect recipe for any theory to deal with any counterexample. Rather, what is also needed is a convincing story of how the revision needed to save the theory you started with from the absurd consequence can be motivated from within your theory and its underlying motivations (2011, 438-439).
It is thus another question whether a proportionalist, Humean view is in fact indefensible. Although both Schroeder and his critics seem to agree that it is, I am more confident that a version of HTR can be salvaged without rejecting proportionalism. If I am wrong, however, then the arguments in this chapter have convinced me that the appropriate course of action is to abandon Humeanism. If one is Humean, then one ought to be proportionalist. And so if the arguments against proportionalist Humeanism are devastating, it is time to abandon HTR.
CHAPTER 4

Over the course of the previous three chapters, I have taken Scanlon’s contractualist moral theory and modified it with a progressively heavier hand. As a result, I have argued: that a contractualist should reject Scanlon’s theory of moral motivation in favor of my own ‘complex’ account (Chapter 1); that one who does so should also replace the ‘primitivism’ about reasons defended in *What We Owe to Each Other* with a Humean account inspired by Mark Schroeder (Chapter 2); and that a Humean ought to be a proportionalist about the weight of reasons (Chapter 3). According to these arguments, an advocate of Scanlon’s moral theory should endorse a proportionalist, Humean contractualism. On this view, most of us have a weighty reason to act permissibly (justifiably), because we have reason to be what Scanlon calls ‘mutual recognizers’, or what I will call ‘justifiers’ – the kind of people who are committed to living with others on terms that they can accept. This latter reason, then, is explainable in terms of our widely shared, strong desire to have intimate relationships with at least some people.1

Since I began this dissertation by accepting Scanlon’s contractualism, the arguments above have convinced me that I ought to endorse the position just described. I will call this revised view The Humean, Contractualist Account of Moral Reasons (HCA), where ‘moral reason’ just means the reason to act according to the criterion of permissibility. HCA stakes out a large, ambitious position, as it is constituted by: a first-order moral theory (contractualism); a

---

1 Of course, I also hold here what I held in the previous chapters – that many of us have a more immediate reason to act justifiably as well, grounded in our desire to live justifiable lives. I focus on the more complex story as a concession to explaining what I think of as the most difficult case, which is an explanation of the reason to be moral for those who don’t care about being able to justify their actions to others.
theory of practical reasons (HTR); an account of moral motivation (*The Motivational Basis of Morality*); and a theory of weighting (proportionalism). Given the scale of the account, then, it would be ridiculous to claim that I am here mounting a complete case for it; rather, I am modifying a well-known view (Scanlon’s contractualism) by adding other, well-developed views (such as Schroeder’s HTR and proportionalism). As a result, the plausibility of HCA rests, in large part, on the independent plausibility of these elements, as well as on the details of how the view is worked out.

I cannot, in what remains of this dissertation, develop and defend every nuance of HCA. I *can*, however, make some initial headway, by clarifying the view, noting what would need to be done in a genuinely full explication of its commitments, and defending it against the most obvious and damning objections. I will begin this process in the current chapter by explaining more fully (and in the most sympathetic way possible) what HCA may actually look like as a positive account, rather than merely as a cobbled together of other views. This will require not merely summarizing the positions laid out in the previous chapters, but also noting those aspects of the view that require further work. Such clarification – although helpful in seeing how the view works – will also render obvious two points: first, that I must, in this project, allow myself to refrain from filling in and arguing for every detail of the view; and second, that HCA is open to many, powerful criticisms. After summarizing HCA, I will, in this chapter, note what seems most obviously in need of further work, and make some gestures at what I would expect the results of such work to be. The remaining chapters of the dissertation, then, will raise and respond to the most serious objections to HCA.

4.1 The Cobbled-Together View
If Scanlon’s intuitions are widely-shared, then the best reason to endorse contractualism (of his preferred flavor) is the manifest plausibility of the idea that impermissible actions are exactly those acts that cannot be justified to others (1998, 155). As discussed in the Introduction, this criterion – taken as a merely formal relation between two sets of actions – need not be rejected by anyone, since the direction of explanation could run either from justifiability to permissibility or permissibility to justifiability. Of course, Scanlon does not intend the merely formal claim, but rather he intends what I have called substantive contractualism, or the view that the justifiability of actions is what explains their permissibility. And so long as he does not smuggle moral concepts into the utilized notion of ‘justifiability’, then, substantive contractualism offers a promising account of the rightness and wrongness of actions. Of course, this is not an obvious nor an easy task, and Scanlon spends a decent number of words providing his own account of ‘justifiability’, which is short-hand for the property of those acts that would be allowed by principles that no one could reasonably reject (1998, 153-155). And so, in a complete defense of HCA, one might want to know whether I will eventually defend substantive contractualism, and if so, whether I will attempt simply to employ Scanlon’s own account of justifiability, or replace it with another, more reductionist-friendly account. I will not, however, provide such a complete theory here, and so we will simply need to keep in mind that one constraint on a complete defense of HCA is that, if it is ever to be an account not just of moral reasons, but of the content of morality as well, then it owes further explanation concerning the contractualist criterion.

The broad goal of HCA, then, is to explain what reason one has when (if) one has a reason to act morally (understood now as justifiably). I have argued that people typically have
multiple reasons to act justifiably, but that our best account of ‘the’ reason to be moral will pick out a consideration that is widely-shared, difficult to evade, and connected properly to other considerations that explain our intuitions about cases in which we want an agent to act from a sense of duty. According to HCA, the moral motivational story is complex, as the widely-held desire for intimate relationships grounds a wide-spread, weighty reason to act justifiably, and it does so by way of explaining one’s reason to be a justifier. In this way, HCA satisfies both desiderata of a moral motivational story, relying on the desire to have intimate relationships to explain the sense in which the reason to be moral seems applicable to nearly everyone, but then relying on the reason (and for some agents, some of the time, immediate reason-grounding desire) to be a justifier, in order to explain the morally relevant reason from which we hope an agent will sometimes act. The final addition of proportionalism to the story, then, utilizes the fact that our moral motivational explanation (at least sometimes) bottoms out in a strong desire. As a result, one’s reason to act rightly tends to be weighty to a degree that typically renders it operative.²

In order to see how HCA actually explains moral reality, let us look at a concrete case.³

Suppose that while on my way to hear a talk by Peter Singer, I walk past a shallow pond in which I see a baby struggling. After a quick look around, I accurately surmise that no one else is around, and that the baby is, in fact, drowning. Without hesitation, I wade into the water, getting

---

² Operative, that is, in a normative sense: one’s reason tends to be weighty to the degree that it is what determines what one has most reason to do. Of course, even when this is the case, the reason may well not be psychologically operative, as one may well not take it to be her weightiest reason, and so may not act in accordance with it.
³ The following example is a modification of Peter Singer’s famous ‘Shallow Water Case’ (1972).
my sandals and pants wet, but saving the child’s life. I then miss the entire lecture while dealing
with authorities, who are attempting to figure out why an infant was alone in a shallow pond.

Rescuing the baby was clearly the right action, and continuing on to the lecture would
have been impermissible, as I could not have justified doing so. Scanlon would say that this is
because some form of the Rescue Principle is among those principles that others could not
reasonably reject (1998, 224), but it is not necessary to employ this explanation of justifiability.
So long as it is overwhelmingly intuitive that walking past the child would have been
unjustifiable, and so acting as I did was the right thing to do, then the case works for present
purposes. Our question, then, is, “What reason did I have to act rightly – that is, to act as I did?”

HCA explains the Shallow Pond case in the following way. Knowing my own
psychology, I can say that I have a strong desire to live a justifiable life, or to be a justifier. In
the example, then, my desire to act justifiably generates an immediate reason to rescue the child,
and this reason is very weighty – in fact, among the weightiest reasons I have on most occasions,
and certainly the weightiest reason applicable in the case. While I enjoy philosophy lectures,
would like to hear Singer talk, am concerned not to flake out on commitments that I have made,
etc., none of these desires are even close to competing with my desire to live a justifiable life.
According to HCA, then, I clearly have most reason to rescue the drowning baby.

However, we can also ask what the situation would have been like if the protagonist were
someone less concerned with being a justifier; what if it were Jake, instead, walking by the pond,
and suppose that Jake did not care one whit about being able to justify his actions to anyone? In
this case, HCA must rely on the full, complex story of moral motivation. Our question for Jake
will be whether he desires intimate friendship or love in his life, and we will assume (and hope!)
that the answer is yes, as most people we meet do have such desires. If so, then the arguments of Chapters 1 and 2 tell us that Jake has a reason to be a justifier grounded in this desire for intimacy, and this mediate reason is what gives him a reason not to walk past the drowning infant. And again, we will assume that his desire for intimacy is not merely one among many, relatively equal desires – but that it, like most people’s desire for love and friendship, is among the strongest desires in his life. And so his reason to rescue the child will not just be any reason, but a weighty (and presumably, his weightiest) reason.⁴

Many questions and challenges arise at this point. Some of them will be addressed in the following two chapters, as I defend HCA against powerful objections. And some of them will simply go unaddressed. However, for now the view itself is still unclear, and so before moving on to objections, I will spend the remainder of the current chapter noting which details most need filling, and what I expect those details might look like. In particular, one might be concerned that HCA has not, in fact, provided a complete account of when one has a weighty, or most, reason to be moral. One reason for this concern is that HCA explains the weightiness of a reason in terms of strength of desire, but ‘strength’ seems to be as much in need of explanation as

⁴ Of course, there is one other variation on the story – the one in which Caligula walks by the pond on his way to ridicule Singer at the lecture. And Caligula, of course, has no moral reason-grounding desires. Instead, all of his desires are for inflicting and watching the infliction of pain and suffering. HCA implies that Caligula has no reason to save the infant, and most reason to walk by (or more plausibly, to sit down and watch with glee). This is a disturbing implication of the view, which has been admitted throughout the dissertation. In the final two chapters, I will say a few things about my take on the Caligula case, as well as why I don’t take it to be devastating to HCA. But none of those comments will deny that, if such a person exists, the normative facts will be just as I have described them: Caligula, if he exists, has no reason to wade into the water to save the infant, and most reason to sit and watch the suffering of the drowning child.
‘weight’. Further, HCA, by itself, does not explain how the total weight of a reason to act is to be determined in a case where there are multiple reasons of varying weight. Both of these issues are difficult, and could be explored with varying levels of care and precision – likely up to the point of dedicating a dissertation to each of them. For present purposes, they do not, however, seem like appropriate topics in which to get bogged down; yes, a complete argument for HCA would need to provide details of how to answer each question, but I have already admitted that this is not a complete argument for HCA. Rather, this is an initial investigation into the viability of the view, and so for now, what seems important is to establish that there are intuitive ways to answer the challenges that do not seem obviously problematic. In the coming sections, then, this is what I will attempt to do.

4.2 Strength of Desires

The issue of *strength* of desire is important because it is so easily understood in a way that cannot be utilized by HCA. In particular, one might think that a desire is strong to the extent that it actually motivates an agent. Thus, if I desire both to take a coffee break, and to finish another section of this chapter, the desire that is strongest is simply the one that has the most motivational force. But such a view is obviously problematic for the following reason: if adopted, HCA would entail that no one ever acts contrary to the force of reason. Since the strongest desire is the most motivationally efficacious one, the weightiest reason is the reason to do the thing that one is most motivated to do. But it seems plausible that one will in fact do whatever one is most motivated to do, and so one will always act in the way favored by the
reasons. On this view, we could actually determine what one has most reason to do by waiting to see what one in fact does!\(^5\)

Another candidate view has strength related to one’s *experience* of the desire’s intensity, such that a strong desire is one that completely occupies my attention, or involves a visceral craving. On such a view, my desires for new motorcycle accessories are strong as a result of the way they direct my attention,\(^6\) whereas my desires to be a loving husband and a careful philosopher are often weak and sometimes non-existent (as they are felt very little or not at all). But this, too, seems problematic, as it seems to me that my husband- and philosophy-related desires are quite strong in an important sense, even if they aren’t always felt. In fact, this second view faces the challenge that *much* of what I take, on introspection, to be my strongest desires are not of the kind that they are regularly ‘felt’ at all. Felt desires are often (although certainly not always) for rather unimportant things, like a new motorcycle accessory, or a big bowl of ice cream.

---

\(^5\) One might object that, whether problematic or not, this seems to be what strength actually means in relation to desire, and so the challenge for HCA is actually to find something other than strength (likely related to strength) that can ground the weight of a reason. However, as I hope will become clear in what follows, I simply think it’s false that the motivational view is what ‘seems right’. To me, at least, what seems more obvious than anything else about the strength of desire is that some of my strongest desires regularly fail to motivate. Many of my strongest desires, like those of being a good husband, father and philosopher can be motivationally trumped by what seem like weak desires – like those that lead me to some distracting activity. While I accept that others may have different intuitions on the most natural account of strength, I take it that there are, in fact, good reasons for thinking that strength can come apart from motivation in the way that I will suggest.

\(^6\) Scanlon has a nice discussion of such desires, which he calls ‘desires in the directed-attention sense’. See further (1998, 39-41).
Both of these problematic accounts of strength seem plausible particularly when one sets out looking for an account in particular ways – in ways that, I think, we should reject. Take the first view discussed above – what I will call the Motivational Efficacy view: this account would seem very natural for someone who arrived at the Humean view via The Classical Argument. That is, if one thought that the best reason to be a Humean is the combination of Reasons Internalism with the Humean Theory of Motivation, then the primary link between reasons and desire would seem to be motivation. On a Classical Humean view, then, the stuff that motivates (desire) is what grounds reasons, and so it seems reasonable that the extent to which something motivates is what is proportional to the weight of reasons. But on the proportionalist model, what is proportional to the weight of a reason just is the strength of desire; and so, a desire is strong to the extent that it actually motivates an agent.

Or consider the second view discussed – what we can call the Phenomenological View. This is the kind of account of strength that one might come to from decontextualized reflection on the question, “what is strength of desire?” After all, we have an intuitive sense of desire as a kind of felt phenomenon, and it seems plausible that strength, applied to a felt phenomenon, would yield the view that the stronger of those phenomena are the ones that are more felt, or that perform their felt function to a greater extent. So examples come readily to mind of the kinds of desires that I sometimes find all-consuming, or that have a powerful bodily component. Perhaps sexual desires, desires for food and drink, or the desire of new parents to get uninterrupted sleep end up being some of the strongest desires on this sort of view.

---

7 The name for this argument comes from Schroeder (2007), and is discussed in Chapter 3, f. 4.
Looking at the motivations for each of these views, however, reveals that HCA will not find either of them compelling. HCA is not moved by The Classical Argument for Humeanism, and so is not committed to a view on which desires are essentially motivating. Similarly, HCA is not impressed by any account on which desire is essentially a felt phenomenon. Rather, HCA is the view that reasons are grounded in desires in the way indicated by its motivating examples. Importantly, the move here isn’t question-begging; I am not arguing that the Motivational Efficacy View and the Phenomenological View fail because they can’t make sense of my account; the argument is that neither of these views is compatible with the arguments for HCA in the first place. That is, the methodology has been to construct a view of reasons that makes sense of paradigmatic Humean examples, such as those used in the early sections of Chapter 3. HCA, then, holds that reasons are grounded in desires in a way that explains why my desire for coffee explains my reason to go to the kitchen, but also in a way that explains why my reason to go to the kitchen is less weighty than my reason to get ready for a pre-scheduled date with my wife. But that means that HCA is committed to an account of strength according to which the desire that explains my reason to get ready for my date is stronger than the desire that explains my reason to walk to the kitchen. This is the case even if the desires that ground the latter reason – suppose, my desires to please my wife and to be a good husband – are felt very little if at all, while the desire for coffee overwhelms my tired brain.

What the coffee/date example shows is that both of the explored models above fail to account for the sense of strength employed by HCA. Strong desires, according to HCA, may fail to motivate, and may not be felt strongly, or at all. And yet there is an intuitive sense in which the language of strength applies. It is, in fact, the case that my desires to please my wife and to
be a good husband are among my strongest desires, in a clear and intuitive sense, even if they fail to motivate me, and even though they often fade into the background of my felt experience. And it seems plausible to me that everyone can come up with examples from their own psychology of desires that seem clearly strong, but that may sometimes fail to motivate. Perhaps these are typically ‘thick’ desires, which are bound up with history, commitments, and which are spread over time; in my own case, this seems to be true, as my strongest desires tend to be concerned with relationships and a form of life. Whatever the explanation, though, so long as there are examples in which one’s strongest desires fail to move her, and fail to be psychologically salient, then the point has been made. But this leaves us with the question: what property, exactly, is HCA pointing to when it claims that a desire is strong, stronger or strongest? There are a variety of views that can make good sense of the intuitive account of strength laid out above. In the remainder of this section, I want to sketch a few different candidate views, and to suggest that a ‘Counterfactual View’ of strength is a promising way to go.

Despite their flaws, the Motivational Efficacy View and the Phenomenological View have a theoretical virtue that a Humean may well find attractive: namely, they are both reductionist-friendly analyses of strength. On a Humean view, reasons are explained in terms of non-normative facts and background desires, making the view promising for anyone who wants to reduce the normative to the non-normative. This advantage would be undermined, however, if desires were features that admitted of measure (as proportionalism holds), and if that measure required normative explanation. Doing so would reintroduce normativity into the explanatory story that sought to eliminate it. On this count, the Motivational Efficacy View and the Phenomenological View fare well, since they explain strength in a perfectly non-normative way.
Desires are made stronger or weaker, on these accounts, by particular, descriptive facts – namely, the extent to which they motivate the agent and their phenomenological salience, respectively.

However, as I have argued, both of these initially intuitive views fail. And so we are left looking for other accounts of strength of desire. What I will suggest here is that the intuitive cases motivating Humeanism and proportionalism also recommend a view of strength that may not obviously be so reductionist friendly. Instead, a careful look at what seems true about strong desires will reveal that there is at least a threat of normativity reentering the picture. However, I will also suggest that a promising view does not do this in a viciously circular way, nor is it clear that strength must admit of normative explanation; it does, however, remain an open question on my view.

As an illustration of how abandoning the clearly reductionist explanations may provide a more intuitive solution, let us look to the other end of the explanatory spectrum – towards a fully normative, non-reductionist explanation. What would such a view look like? One candidate view would simply reintroduce reasons into the explanation, such that a desire is strong to the extent that one has reason to want or to pursue the desired end; or perhaps one desire is stronger than another if there is reason to pursue the first over the second. Such a view would, in fact, help with our intuitive cases, as we could say that strong desires aren’t always those that motivate, or those that are salient, since we might have little or no reason to pursue the motivating or salient desire. But of course this way of explaining our judgments is circular: reasons are explained by desires, which have strength, which is explained by reasons, which are explained by desires, etc. In addition, while separating so starkly the strength of desires from one’s psychology helps intuitively in the sense of allowing that strength might depart from
motivation and phenomenological salience, it errs in the other direction: on such a view, strength of desire ends up being about what one has reason to desire, rather than what one actually desires. Although the fully reductionistic explanations of strength seem to tie it too closely to particular, psychological features, a radically normative explanation too far removes strength from facts about one’s psychology.

Fortunately, we need not adopt the circular explanation above to make an advance over the Motivational Efficacy View and the Phenomenological View. And in fact, it is not the most natural explanation. Instead, what seems true about the correct view of strength is that it is a counterfactual view. As I previewed above, I will not here argue for a particular account, but I will sketch the intuitive argument for this way of conceiving strength, and in the following sections, provide some minimal conditions on such a view.

Let us return to the toy case of comparing my desire for coffee and my desire to be a good husband. These desires ground different reasons, and so we will want to be able to compare their strengths. Suppose that in some particular case, I am trying to decide whether to do something important for my wife, or whether to skip that activity in favor of some much-needed coffee. I haven’t slept much in days, and I can think of nothing other than wonderful, dark, acidic coffee. So I begin to blow off my wife’s requested activity in favor of a coffee run, when a friend sees me and asks what I’m doing. I explain to her that I want coffee much more than I want to be a good husband at this moment, to which she replies, ‘no you don’t’. We can imagine my close friend saying to me something like, “Look. I know you are really tired, and you aren’t thinking straight. But you don’t actually want coffee more than you want to be a good husband. If you were clear-headed, and properly thinking about your wife and coffee, and had
all of the relevant information, you certainly would not endorse your desire for coffee over your desire to be a good husband. Surely, in more ideal circumstances, if the two came into conflict, you would pursue being a good husband over getting coffee!" 

My wise friend here seems to be on to something. When a desire is strong, what seems true is that it is a desire that I would endorse, or perhaps one that I would pursue, if I were thinking about it properly (whatever that turns out to mean). So while I do, in fact, love good coffee, I also recognize (now, in a relatively ideal deliberative setting) that it is not all that

---

8 One might claim that my friend doesn’t sound like she is explaining which of my desires is strongest, but rather which of my desires I wish was my strongest. Of course, if one is already committed to the Motivational Efficacy or Phenomonological Views, then this is exactly how such a case will sound. But I have tried to explain why I am not moved by such views. Having given the motivating examples for the view, I do not think that I can say much more to convince such an opponent that this is a plausible way to think of strength. It seems natural to me to think that my friend is, in fact, helping me to figure out which of my desires is in fact strongest, despite what is salient or what is moving me.

9 It’s worth noting all of the ways in which this (intuitive, I’m suggesting) claim is vague: it mentions both endorsing and pursuing, and it leaves it open as to what ‘thinking about it properly’ entails. At this level of abstraction, then, the counterfactual condition could concern either a judgment (endorsing) or an action (pursuing), and could require a wide variety of deliberative constraints. Although I will not, in this dissertation, take a stance on any of these ambiguities, it may be helpful to note that, even at this level of abstraction, the counterfactual view may be analogous to Gauthier’s counterfactual view on deriving utility from preferences. According to Gauthier, one rationally ought to maximize her value, or utility, and value is the measure of ‘considered preference’. Preferences are ‘considered’, then, ‘if and only if there is no conflict between their behavioural and attitudinal dimensions and they are stable under experience and reflection’ (1986, 33-34). Interestingly, Gauthier accepts this view over the economists’ ‘revealed preference’ view, according to which what one prefers is inferred from what one actually chooses. This rejection, it seems to me, is analogous to my rejection of the Motivational Efficacy View above: just as I argued that desires are not necessarily strong because they motivate us, Gauthier argues that one does not necessarily prefer something just because she chose it. The connection between our positions gives me some confidence that Gauthier is an ally in developing an account of the psychological grounds of normativity (reasons for me, value for him) that is not completely reductionistic as a result of incorporating counterfactual constraints.
important to me. Given few constraints or competition, I’ll drink good coffee; but that desire is not one that I would endorse over many other, sometimes less-salient desires. On the here-investigated ‘Counterfactual View’, if my desire for coffee isn’t one that I would endorse or pursue in some ideal setting, then my desire for coffee isn’t all that strong – or at least, it is much less strong than many of my other desires. On the other hand, desires like being a good husband, a good father, a careful philosopher, and living a justifiable life are ones that I readily endorse, and would pursue over most other desires. And this diagnosis seems to get things right. Coffee sorts of desires are not all that strong when compared to family, career and morality desires. These latter desires, even if they do not occur to me all that often in my conscious experience, are paradigmatically strong desires.

A Counterfactual View occupies a place somewhere between the obviously reductionist views and the obviously non-reductionist views on the spectrum of the normativity of explanation. The entrance of idealizing conditions makes it seem plausible that the Counterfactual View is not completely reductionist-friendly. It may sound a bit too close to the view that strong desires are the ones that it would be rational to pursue, or that one ought (in some sense) to pursue. And this is clearly a normative explanation. But such a summary is not necessitated. One might try instead to articulate an identity-based view, on which what one would endorse or pursue on reflection is determined by facts about one’s practical identity.\textsuperscript{10} While such a view might constitute a normative explanation, it’s not obvious that it must; whether it does depends on whether the theory of practical identity could be cashed out in a

\textsuperscript{10} That is, one might take inspiration from Korsgaard’s model of the source of normativity (1996).
reductionist-friendly way. Alternatively, one might pursue a hierarchical model of desire, interpreting the idealizing condition as being what one desires to desire.\textsuperscript{11} While such a view faces challenges, at least one philosopher seems to think that such a view need not bottom out into normative explanation (Frankfurt 1987).

A Counterfactual View, schematic as it is in presented form, does not provide a particular explanation of the idealizing conditions; further, it does not even require that such conditions constitute an explanation or analysis of strength. And this is yet another way in which the intuitive diagnosis made by the Counterfactual View does not obviously render it non-reductionist. All that is required by the view is that certain conditions hold of the correct account, whatever it is. And this, minimal, strategy is the one that I will pursue in the following section. Since I do not want to take a stand on the thorny issue of whether HCA must adopt a non-reductionist foundation – so long as I can avoid it – I will here argue only what I must in order to understand HCA sufficiently for purposes of evaluation. And for that, all we need to know is what conditions are sufficient for rendering some desires strong or stronger than some others. I thus suggest minimal, sufficient, counterfactual conditions that seem plausible regardless of whatever else is true, and leave the question of what ultimately grounds these conditions for another project.

4.3 Strong, Stronger and Strongest Desires

\textsuperscript{11} And in fact, such a view seems naturally to lend itself to an explanation of the strength of desire, since Frankfurt intended his view as an account of one’s identity. If a desire, then, is one that I desire to have, it is stronger than one that I desire not to have by virtue of being a desire that forms part of who I am. Utilized by the Counterfactual View, then: the strength of a desire that \( p \) depends on how much one would desire that \( p \) if one had the desires that one desires to have. See further Frankfurt (1971; 1987).
If some version of the Counterfactual View is true, then we know that, when one desire is stronger than another, it must be the case that the desiring agent would, under some idealized condition, endorse or pursue the stronger over the weaker. What we do not know is whether the stronger desire is stronger because of this feature, or whether some other explanation is responsible for both facts. For present purposes, I won’t worry about this question, but will instead be content to try to articulate some form of the counterfactual above that serves as a plausible sufficient condition for the correct view of strength.

However, the initial counterfactual articulated above is not yet a plausible sufficient condition. To show why it is not, consider the following example:\textsuperscript{12} I, in a possible world very close to this one, have as one of my (intuitively) strongest desires the desire to please my wife. I also regularly desire ice cream. It seems, in this world, as it does in the actual world, that the correct theory of strength should entail that my desire to please my wife is stronger than my desire for ice cream. However, after spending two weeks on a special vacation with my wife, during which time I catered to her every whim – and during which time I abstained from ice cream, because the sight of it makes her nauseous – I find myself really wanting ice cream. Now the question arises: in this case, even if I am in a suitably idealized deliberative environment, would I necessarily endorse or choose to promote my desire to please my wife over my desire for ice cream? Perhaps not. One might think that even a good husband would allow himself an idiosyncratic rank-ordering of desires in this case. It may be, then, that in this case I would (even ideally) endorse or promote my desire for ice cream over my desire to please my wife. But

\textsuperscript{12}Thanks to Henry Richardson for suggesting this model of counterexample.
doing so does not seem to entail that my desire for ice cream is stronger than my desire to please my wife; it seems more natural to think that, though my desire to please my wife is stronger, it is reasonable in this case to privilege the weaker desire. So for the counterfactual condition to work, there must be an additional feature built in. Since desire X can be stronger than desire Y even if I would, very occasionally, endorse or pursue the satisfaction of Y over X, the condition above needs to be weakened. Perhaps desire X is stronger than desire Y if one would tend to or would generally endorse X over Y in some idealized conditions.

The language of ‘tending to’ or ‘generally’ endorsing is perhaps a bit too vague, however, and we can avoid this vagueness by weakening the condition just a bit further. The previous case of pleasing my wife vs. getting coffee seemed most plausible when we assumed that I would always endorse or pursue the former over the latter, and what the ice cream counterexample brings to light is that we plausibly do not change our judgment about strength even when I would almost always endorse or promote the one over the other. And while ‘almost always’ is a bit vague as well, we can at least understand that if I would almost always do X, I would do it nearly all of the time – that doing not-X would be a rarity and the exception, and that it would be the result of a very non-standard set of circumstances.

Given these considerations, I suggest

**Stronger**: a desire X is stronger than a desire Y if the desiring agent would, under some idealized condition, always, or nearly always, endorse the having of or pursue the satisfaction of X over Y.

Of course, much of this claim is in need of further explanation, as I have not provided the details of the exact idealizing conditions, and I have admitted that the language of ‘almost always’ is
imprecise. Despite these limitations, however, Stronger seems overwhelmingly plausible. Whatever else is true about the strength of desires, it seems true that when a pair of desires meets this condition, the first is stronger than the second. If so, then we might formulate an additional condition that will be helpful when evaluating HCA:

**Strong**: a desire is strong if the desiring agent would, under some idealized condition, always, or nearly always, endorse the having of it or pursue the satisfaction of it over the vast majority of other, competing desires.

This, of course, is just to say that a desire is Strong if it is Stronger than the vast majority of other, competing desires, which seems like a natural claim. In fact, it is likely weaker than what is in fact the case, as it is plausible that some desires that are intuitively strong need not be stronger than quite so many other desires, so long as it is stronger than desires meeting some contextually-determined standard.\(^{13}\) However, it seems plausible that all desires that meet the criterion set by Strong are intuitively strong, even if there are additional, intuitively strong desires as well.

Finally, then, a natural extension of Stronger is

**Strongest**: a desire is strongest, among some group of desires, if the desiring agent would, under some idealized condition, always, or nearly always, endorse or pursue it over any other desires in that group.\(^{14}\)

\(^{13}\)Thanks to Steve Kuhn for this suggestion.

\(^{14}\)The addition of ‘groups’ of desires in this condition is so as to provide an understanding of strongest that allows for more than one strongest desire. Perhaps I have an entire class of desires that are my strongest; Strongest allows that this may be the case by showing how a desire can be strongest among some group. Of course, one could name all of the desires of some agent as a ‘group’ in order to discuss the possibility of a strongest desire simpliciter as well.
And this claim, of course, is equivalent to the claim that a desire is **Strongest** if it is **Stronger** than all other desires in some group. These latter two conditions will be helpful when combined with proportionalism, as they will allow us to move beyond the judgment that some reason is weightier than another to the judgments that some reason is weighty or weightiest.

If **Stronger**, **Strong**, and **Strongest** are true, then we will be able to explain at least some instances in which a reason is weightier than another, weighty, or even weightiest. Knowing my own psychology, for instance, I am confident that my desire to live a justifiable life is stronger than my desires for very many other things. As a result, I know that I have one reason to act justifiably that is weightier than, say, one reason to steal good coffee – and in fact, I have one reason to act justifiably that is simply **weighty**. In very many discrete instances, my desire to live a justifiable life will be stronger than any competing desire (that is, any desire in the group that is competing to motivate me), and so my reason to act justifiably will, in fact, be among my **weightiest** reasons.\(^{15}\)

However, these judgments represent only a small, simplistic aspect of the relevant normative space. Providing strength conditions for desire will allow us to ask more, difficult questions about how desires interact with one’s ability to promote them to generate weight, as well as how the weights of reasons add up. Unfortunately, the detail and the difficulty of these questions explodes at this point, and so all that I can do in what remains of the chapter is note

---

\(^{15}\) This simplistic example leaves aside the second proportionality of proportionalism – that the weight of one’s reason to phi is also proportional to the degree that phi-ing promotes the grounding desire. For the sake of simply seeing how a Counterfactual View combines with proportionalism, then, we can read the case as holding the promotion relation fixed at ‘maximally promotes’ for all relevant desires.
what a view like HCA requires, and that there doesn’t seem to be any reason to assume that it
cannot be provided.

4.4 Numerical Weights

Recall that HCA’s theory of weight – proportionalism – actually involves two
proportionalities: the weight of a reason to phi is proportional, not just to the strength of the
grounding desire, but to the degree to which phi-ing promotes the grounding desire. Neither this
general formulation of proportionalism, nor anything I’ve said here, explains how these two
proportionalities interact in determining particular weights. All we know is that some formula
like the following is true:

\[ W_t = (S_t) \times (P_{\text{rom}}) \]

where \( W_t \) is weight, \( S_t \) is strength of a reason, \( P_{\text{rom}} \) is the degree of promotion, and \( \times \) is a
‘product’ notation. The weight of a reason, then, is equal to the product of the strength of the
grounding desire and the degree to which the recommended action promotes the grounding
desire.

Note that there are at least two difficult questions about understanding proportionalism in
this way. First, it looks like a mathematical formula, which would imply that reasons might have
actual numerical weights; but could that be right? Second, depending on the answer to the first
question, how are the weights of multiple reasons added or compared? If we do not accept a
numerical understanding of the proportionalist formula, is there an intuitive way to understand
how my reason to act justifiably (grounded in my desire to live justifiably) and my reason to act
justifiably (grounded in my desire for intimate relationships) combine to form an even weightier
reason?
The first question is raised by Schroeder, who seems to think that the very talk of reasons ‘having’ weight implies a numerical model, and that the notion of numerical weight is problematic. His reasoning is as follows:

I would be strongly hesitant about reifying weights of reasons. What are those weights? Are they numbers? As far as I can tell, our practice of weighing reasons only supports the view that reasons can sometimes weigh more or less than one another – I see little evidence that they have numerical weights, and much evidence that they don’t. It seems at least conceivable, for example, that there are pairs of conflicting reasons, neither of which weighs more than the other (2007a, 101 f. 27).

I do not find Schroeder’s worries here very compelling. One piece of evidence that reasons have weight (and are not simply ‘weightier than’ or ‘less weighty than’) is that it seems true that some of my reasons are simply weighty. Even if we could somehow drum up a case in which my reason to rescue a drowning infant was relatively less weighty than many other of my reasons, this does not exhaust our observations about its weight. My reason to rescue the infant is very weighty, even if it is relatively less weighty than others. Of course, Schroeder could try to explain these intuitions with a comparative model, but my point is only a prima facie one, which is that there is, in fact, reason to find it natural to talk of weight, simpliciter.

Schroeder does then claim that there is much evidence that reasons do not have numerical weight, which he seems to think is the only way to account for weight. His example of such evidence is that it seems at least conceivable that there are conflicting weights, neither of which is weightier than the other (and presumably, which do not weigh precisely the same). However, the conceivability of such a scenario does not imply that it’s true. It might be conceivable, but if
the best theory of weighting entailed that there are not, in fact, such scenarios, then we might judge that the scenario is conceivable but false.

I am not here taking a stand on whether we should expect to be able to use the proportionalist formula numerically or not; rather, I am simply noting that I am not convinced that doing so is a clearly bad idea. It seems to me a live question as to whether we should expect a mathematical interpretation of the proportionalist formula to be defensible. However, even if it turns out not to be, I am also not convinced that this would be terrible for the Humean. Perhaps, Schroeder is right to worry about numerical weights, but that it’s also the case that the proportionalist formula doesn’t require it. Perhaps, that is, the proportionalist formula should be read as saying only that, if the promotion relation is held equal, stronger desires generate weightier reasons; and that if the strength of desire is held equal, that a stronger promotion relation generates a weightier desire. This would still be useful in working through cases in which we had an intuitive sense of a desire’s relative strength and the degree to which acting promotes that desire.

4.5 Adding Reasons

The second question raised by the proportionalist formula concerns the addability of reasons. HCA, like any Humean view, intends for reasons to be addable, such that multiple reasons of some (relative or not) weight add up to equal a weightier reason. This is one reason for hoping that a numerical account might be worked out, as it would provide a straightforward explanation of how reasons are so addable. However, there is an additional worry about the numerical view of weight here, as straightforward mathematical addition sometimes seems like the wrong model for adding reasons. Very, very many reasons of very, very little weight ought
not, we might think, to add up to equal a very weighty reason. For instance, millions of reasons to prevent mild, 10-minute headaches ought not to add up to outweigh a reason not to kill someone. So we might, yet again, find ourselves unclear on the issue of whether reasons having numerical weight would be helpful or not.

Even more serious, I think, is a concern about addability raised by John Horty that does not challenge merely numerical models of weight. In his (2012), Horty raises an objection to ‘reason amalgamation’, or what I have been calling their addability. The worry is that, although it is natural to think of two atomic reasons that their independent weights might be added together to generate a combined, weightier reason, there are cases in which this model doesn’t seem to make any sense. I will call such cases ‘Amalgamation Counterexamples’.

Horty’s central Amalgamation Counterexample concerns the reasons that one has to go for a run. It seems plausible, he suggests, that the following case obtains: the fact that it is hot is a reason not to run; the fact that it is rainy is a reason not to run; but the fact that it is hot and rainy is not a stronger reason not to run – in fact, it is a reason to run, as hot, rainy runs are refreshing (Horty 2012, 2.2.3). And indeed, at first glance, this looks like a real problem for HCA. It seems plausible that one could desire not to run in the heat, generating H, and that one could desire not to run in the rain, generating R. But given that H and R are two reasons to refrain from running, proportionalism says that each will have some weight; and so if they are ‘addable’, we would expect it to be the case that on has the reason HR, equal in weight to the combined weight of H and R. But instead, since hot, rainy runs are refreshing (or at least, since

---

16 With special thanks to Steve Kuhn for raising this objection, and to Henry Richardson for helpful discussion concerning how the proportionalist ought to respond.
one might think so!), HR in fact is a reason *to go* running. In this case, it looks like H and R fail to amalgamate, or to add up to a reason equal in weight to the combination of their respective weights.

Horty’s objection is interesting and powerful. I take it to be a strength of proportionalist Humeanism that each reason has weight, precisely because we do seem to amalgamate reasons, and the idea of atomic weights lends itself to this practice. If the Amalgamation Counterexamples turn out to require HCA to say that not all atomic reasons to do the same thing are addable, then this will likely seem to rob HCA of one of its potential strengths.

It may also seem that the Amalgamation Counterexamples pressure HCA back towards a picture like Schroeder’s, as he has avoided the issue of adding atomic reasons by making sets of reasons the bearer of weights. And indeed, I can see the attractiveness of trying to incorporate the idea that sets have weight rather than individual reasons having weight, and it would be interesting to investigate whether such a move could be made by a proportionalist, rather than by one who endorses Schroeder’s Recursive Account. However, as Horty points out, it is not clear how helpful the bare move to sets is, since it simply relocates the problem: Schroeder still requires that ‘addable sets’ amalgamate, but that means that he must be able to explain how a set of reasons composed of two reasons is weighted as it is as a result of the two singleton sets composed of those reasons (2012, f. 17). If \{H\} and \{R\} are singleton sets composed of the reasons introduced in Horty’s example, then Schroeder faces the same challenge that HCA faces: namely, explaining the weight of \{H, R\}, given that \{H\} and \{R\} are addable sets, but ought to add up to a reason *not to go running*, whereas \{H, R\} is a set of reasons *to go running*. 

150
Moving from atomic reasons to sets, then, may in fact be promising, but it’s not clear that one could simply adopt Schroeder’s version of such a view and thereby avoid Hory’s challenge. Further, since HCA is proportionalist, which Schroeder’s view is not, developing a view in which sets of reasons are the weight-bearers that is HCA-friendly would be a not-insubstantial (although interesting) project. For these reasons, I will not here attempt to avoid Hory’s counterexamples by moving to sets, even though I see the appeal in so moving. Instead, in the remainder of this section, I will offer what I take to be the best model for responding to the Amalgamation Counterexamples for a proportionalist Humean view that assigns weight to atomic reasons.

The first thing to note is that Hory’s challenge is not an objection to HCA; indeed, it is not even an objection to proportionalism or Humeanisms. It is an objection to all theories of reasons that take multiple reasons for the same action to be addable – that is to say, it is an objection to virtually every theory of reasons I have encountered. On the one hand, this is a bit of a companions-in-guilt defense: by acknowledging the scope of the worry, I avoid the appearance that there is something particularly worrisome about HCA. I, other Humeans (including Schroeder, if what I wrote just above is true), constructivists, and realists about reasons, all must face Hory’s worry. Such an observation doesn’t make the worry less worrisome; but it does keep it from placing HCA at a disadvantage relative to other theories of reasons.

The ways in which different kinds of theories respond to the Amalgamation Counterexamples will likely differ, however, which brings me to my second point. HCA is a Humean theory, and so I will suggest that its response to the counterexamples must utilize the
Humean explanation of how reasons are explained and individuated. In short, I think that the Humean is able to diagnose a sloppiness in individuating reasons that occurs in the most troubling counterexamples. However, I also admit that making this explanation will reveal a potentially odd consequence of Humean views.

Let’s reconsider Horty’s case. The problem for HCA is that reasons are just facts grounded by a background desire, so as long as someone has the desires that ground the reasons stipulated by Horty, then the counterexample goes through. And it seems, at least at first, as though such desires would be common. Sinem, for instance, does not want to run in the blistering hot, and does not want to run in the rain; she does, however, want to run when it’s blisteringly hot with a refreshing rain. And this desire set generates Horty’s counterexample.

In describing Horty’s case with an eye towards Humean explanation, however, a diagnosis practically suggests itself. Since I had to come up with the desires that ground the relevant reasons, I was forced to fill in a few details of Sinem’s psychology. Desires explain reasons, so when we have different reasons, we have different desires. Why might it be true that Sinem has the reasons she does? Well, because something is relevantly different about running in the heat and rain, and my desire relevant to that conjunction is different than the desires relevant to each of the conjuncts. But that means that it’s not plausibly true about Sinem that she doesn’t want to run in the heat and that she doesn’t want to run in the rain; after all, she wants to run in the heat and rain. So what’s more likely true of her is that she doesn’t want to overheat or to get chilled, and that these more basic desires explain why she thereby does not want to run when it’s either just hot or just rainy. When it is hot and rainy, however, neither of those desires
generates a relevant reason not to run; and in fact, because Sinem wants to run when there is a rain that refreshes from the heat, she in fact has a reason to go running in the heat and rain.

This seems like a very natural, plausible story for a Humean to tell. And in fact, some explanation like this will be necessary for dealing with all sorts of interesting cases that arise from conjunctions of states of affairs. Consider a case about which HCA will need to say something similar: I may desire milk in my tea, and lemon in my tea; but I desire that there not be both milk and lemon in my tea. On the surface, this case looks to pose a different, but related difficulty for HCA, as it appears to suggest that reasons to do two, compatible things do not add up to a reason to do the conjunction of those things. But a diagnosis like that in Horty’s case seems even more obvious in this one, as the case comes off as being sloppily described. Although I would likely never think to say it, I do not have a desire for lemon in my tea, or for milk in my tea; I have a desire for lemon if no milk, and milk if no lemon. Or rather, in fact, I probably have a preference between the two, such that I have a desire for milk in my tea, but lemon if no milk. The reasons in this case, then, are unmysteriously not addable. I have a reason to put milk in my tea, and a reason to put lemon if there is no milk. Getting clear on the actual content of one’s desires will, it seems to me, often reveal that the challenge of such cases is illusory.

The above diagnosis is, I contend, natural and plausible. In addition, because Humeanism explains the agent’s reasons by reference to desires, this move is not ad hoc. Desires are real, and they are some way or another. My suggestion is that what makes these

---

17 Thanks to Henry Richardson for this case.
cases seem real is that normal people may have desires that are a particular way; and on understanding the agents’ actual desires, the mystery of the Amalgamation Counterexample disappears. However, there is an additional worry that might be raised at this point: since it is, in fact, the case that desires are some way or another, and which way they are is an empirical matter, I cannot stipulate that agents in cases like those described have the desires described. And this is true. I cannot stipulate the case, and so have instead relied on the intuitive plausibility of the diagnosis. It seems likely to me that the cases are as I have described rather than some other way. But I have to admit that, real psychologies being idiosyncratic as they are, the cases could be different.

What might be the case, in Horty’s example then, is that Sinem just has the following, very strange desire set: she really doesn’t want to run in the heat; she really doesn’t want to run in the rain, but she would quite like to run in the heat and rain. This desire set, then, cannot be further explained in a way that allows us to rationalize it. Sinem is not here, as she was in the other case, desiring different features of the heat and rain. She has, in this case, what seems to be a paradigm instance of an irrational desire set.

If Sinem’s revised case is an objection to HCA, it must run something like this: HCA claims that reasons are addable, but sometimes the addition of atomic desires will end up being weighed against the conjunction of those atomic desires, but with the desire set on each side bearing different weight. Sinem’s case, for instance, will look like this:

\[
((H) \& (R)) \neq (H \& R).
\]

And surely, the objection might go, that doesn’t make any sense. If H and R are addable, then the conjunction of H & R must equal H & R. And I agree. The case doesn’t make any sense.
But that also doesn’t seem surprising to me, since Sinem’s desire set doesn’t make any sense. It is, as I said, a paradigm instance of an irrational desire set. Is it the job of a Humean theory to understand this case in a way that makes sense (given that the agent’s psychology doesn’t make sense)?

I can see the merits of answering this question in both ways. On the one hand, Humeanism is a view that ties the normativity of reasons to the actual psychology of agents, and so when the psychology is messy, the normative reality is messy. In Sinem’s case: when the agent’s psychology doesn’t make any sense, the normative situation one is left with doesn’t make any sense. Of course, there are questions for one who goes this route, such as: is there a true claim of the form, “Sinem has most reason to phi”? And I do not have any intuitions about whether we should expect that.

On the other hand, HCA might admit that messy psychologies sometimes imply messy normative situations without giving up on the idea that the correct view ought to be able to make some progress against the ‘mess’. For instance: HCA has, just in this chapter, introduced an idealizing condition on the notion of ‘strength’ of desire. Might this not help to determine something more helpful to say about Sinem’s situation? After all, she has, in some sense, conflicting desires: her desire to run in the hot rain conflicts with her desires not to run in the heat and not to run in the rain. These desires have strength, and so an account of strength will determine how they relate to one another. The kinds of idealizing conditions (vaguely) introduced here are exactly the kind that would root out irrationalities in desire sets. After all, one condition on a desire’s being stronger than others is that it be a desire that would, under some idealized conditions, be endorsed or its satisfaction pursued over those other desires. But it
is difficult to imagine what idealized conditions HCA might ultimately appeal to that would allow Sinem to retain her irrational desire set. If she were, say, reflecting on the desires, how the relevant states of affairs interact, had all the relevant information, etc, surely, we might think, she would be forced to come to the conclusion that heat and rain are only bad-making features of a run when they are not conjoined (or, alternatively, that the conjunction of heat and rain is not a good-making feature of a run). Although such a move seems initially promising, since I will not here be able to deliver a complete argument on its behalf, I offer it only as one possible way for a view such as HCA to respond to the problem of irrational desire sets.

Consideration of the topics of the proportionalist formula and the addition and comparison of weights thus raises many, very difficult questions. In a complete argument for HCA, these questions are arguably among the most pressing that I would need to answer. However, in providing the initial sketch of a new theory of moral reasons, I cannot take the time to answer them here. I have noted that, while there are challenges to be addressed in providing any particular account, none of them seem, from the outset, to be insurmountable. And so I will not, in the remainder of the dissertation, commit to any particular details of how a proportionalist account of weighting is to be worked out. Instead, I will simply rely on an intuitive understanding of proportionalism, alongside the assumption that reasons, in some way, can be aggregated and compared.

4.6 The Complete Picture: HCA in Action

At this point, I have completed a sketch of HCA, having argued for the primary aspects of the view, but having also noted details that would require further attention in a full explication. According to HCA, moral reasons are reasons to act justifiably, and these reasons are explained
by the desires of particular agents. As a result, idiosyncratic agents will have all sorts of idiosyncratic reasons related to moral action, but HCA points to the immediate desire to live justifiably, as well as the desire to have intimate relationships – which grounds a mediating reason to live justifiably – as providing promising explanations of many actual agents’ moral reasons. Further, we can now see that, given the strength of these desires for many of us, moral reasons will often be *weighty* reasons (so long as moral action promotes those desires to a sufficiently high degree); in fact, if endorsing some other desire over one’s desires to live justifiably and to have intimate relationships is nearly unthinkable for many of us (as it is for me), then in many situations, it may even be the case that an agent will have *most reason* to act justifiably. This would be an intuitive result.

Mark Schroeder makes the point that “adding Proportionalism to the Humean Theory of Reasons yields the thesis that you always ought to do that action which best promotes your desires on balance” (2007, 98). And while I have not, in this dissertation, accepted all of the claims that allow Schroeder to conclude so specifically, it does seem true that HCA (assuming the addability of reasons) entails a similar thesis: that one has *most reason* to do that action which best promotes one’s desires. Since the weight of a reason is proportional to the strength of the grounding desire and the degree to which the recommended action promotes that desire, and since reasons can be aggregated (such that multiple reasons to phi result in a *weightier* reason to phi), what one has *most reason* to do is a function of what *most promotes* one’s overall *strongest* desires. The summary of having ‘most reason to do what best promotes’ one’s desires thus does not seem inaccurate. However, despite Schroeder’s point sounding like a criticism of the view, such a summary is compatible with the fact of moral reasons (that many of us have them, and
that they are weighty, or even weightiest) depending on how many of us have the moral reason-
grounding desires thus far discussed.

What this summary of HCA suggests is that it shares some important similarities with an
extant class of moral theories – namely, Hobbesian theories, similar in varying ways to both the
classic exposition by Thomas Hobbes himself (1968), and to the leading contemporary
Hobbesian, David Gauthier (1986). According to Hobbes, the laws of nature command each
individual to survive, whether that means seeking peace and covenancing with others, or
preemptively attacking one’s competitors (1968). As a result, a popular interpretation of Hobbes
(although there are very many others) is that he is an ethical egoist, and so on his view what one
ought to do is whatever is in one’s interest. Gauthier, then, adopts the Hobbesian label, but
begins from the different claim that what is rational for one to do is what maximizes the
satisfaction of her considered preferences (1986). And while such a focus on one’s ‘preferences’
is different from Hobbes’ account of survival or interest, the two views share the grounding
claim that what one (rationally or morally) ought to do is determined by ‘considerations of one’s
self’ – whether those be considerations of one’s interest, or consideration of one’s preferences.
Further: both Hobbes and Gauthier believe that such a claim is compatible with the existence of
morality because considerations of one’s self will, in our world at least, entail that one ought to
act according to a contractarian account of moral rules.

---

18 In particular, one of the most influential commentators on Hobbes calls his view a form of
‘rule egoism’, as it commands that one follow the rules of one’s covenant for the sake of
promoting her interest (Kavka 1986).
It is likely obvious that HCA shares these same basic features with Hobbes and Gauthier: it holds that what one has most reason to do is determined by considerations of one’s self (in this case: one’s desires), and that this is compatible with the existence of morality because such considerations generate (for many of us) most reason to act according to the contractualist criterion of permissibility. HCA, then, has the general structure of a Hobbesian theory.

Unfortunately for HCA, however, this connection isn’t immediately helpful. Hobbesian views tend not to be very popular, as there is much skepticism that consideration of one’s own interests, preferences or desires can in fact generate the existence of wide-spread, weighty moral reasons. Some of this doubt, I imagine, comes from skepticism that people’s strongest desires actually have any connection to moral action; if this is the case, then the arguments of Chapter 1 have claimed to advance the Hobbesian cause. Counter-intuitive or not, if my Scanlon-inspired argument concerning intimacy is plausible, then a very strong, wide-spread desire (for intimacy) in fact grounds a moral reason.

There is another, famous reason to be skeptical of the Hobbesian project, however, and that is the problem – raised and named by Hobbes himself – of the Fool. And this, I take it, is one of the best reasons to reject a Hobbesian view, if no solution is forthcoming. As I now move into the defensive portion of the dissertation, I will take Hobbes’ Problem of the Fool to be the first objection that must be met. So meeting it is the goal of Chapter 5. In Chapter 6, I will then pull together and respond to a collection of objections that HCA is, in various ways, unsatisfying or extensionally incorrect. I will then conclude in Chapter 7 by evaluating how successful these responses have been, and to what degree we should see the adoption of HCA as ‘costly’. 
CHAPTER 5

In the conclusion of the previous chapter, I noted that the Humean Contractualist Account of Moral Reasons (HCA) sketched in this dissertation takes the rough form of a Hobbesian view. This is so, not because it aims to make good on all of the substantive claims made by Hobbes,¹ but because HCA claims to generate reasons to act according to a contractualist criterion of permissibility from considerations of the agent’s own desires. The challenge to be addressed in the current chapter is that Hobbesian views – even if they can be made initially plausible (which many doubt) – are subject to the very difficult ‘Problem of the Fool’. According to this challenge, considerations of one’s desires cannot generate stable reasons to act morally, since it is at least possible (and probably actual) for it to be the case that one’s desires are best promoted by acting immorally. It thus looks like HCA may have inherited a very difficult problem.

In fact, I believe that the initial Problem of the Fool can be met relatively easily, in roughly the same way that Hobbes suggested and that Gauthier formalized. In short, their response to the Fool was to show that it best promotes one’s desires (or for Gauthier, one’s ‘preferences’), not to act rightly, but to be a ‘moral person’ – that is, a ‘covenant keeper’, or what Gauthier calls a ‘constrained maximizer’. In the first half of this chapter, I will make my own,

¹ Indeed, HCA does not so claim. Simply by virtue of adopting a desire-based theory of reasons, as opposed to discussing ‘self-interest’, I have already left behind any kind of pure Hobbesianism. Further, I will admit below (and again in the following chapters) that morality does, likely, sometimes conflict with promoting one’s desires, whereas Hobbes’ goal in responding to the Fool is precisely to deny that morality ever conflicts with one’s interest. HCA, then, is ‘broadly Hobbesian’, or ‘takes the rough form of Hobbesianism’, in that it seeks to generate moral reasons from considerations of what one desires, by referencing the benefit to oneself of acting rightly. As a result, HCA is sufficiently Hobbesian to inherit the problems of the view.
Hobbesian argument of this form. However, the game-theoretic discussion to emerge in the wake of Gauthier’s initial response has shown that the Problem of the Fool is only the beginning. If we accept the initial Hobbes-Gauthier move in response to the Fool, then we are met with a new, more difficult challenge. I will call this second challenge the Disposition-Act Problem (DA Problem), as it points out that a reason to adopt some disposition (that of a ‘covenant-keeper’, let’s say) is not a reason to act in any particular way (to keep one’s covenant in any one instance). So in attempting to solve the Problem of the Fool, the Hobbesian has failed to generate the intended conclusion of his view – namely, that one has reason (or *weighty* reason, or *most* reason) to act morally.

In the second half of this chapter, I will accomplish two tasks: first, I will show that Fool-DA Problems are not unique (in form) to contractualism, and that in fact, other normative theories have come up with interesting, partial solutions to various versions of these problems. Second, then, I will draw on these partial solutions in building my own complete solution to the Hobbesian’s Fool-DA Problems. I will then conclude by showing how my specifically Hobbesian solution may be adopted by other positions – in particular, by indirect consequentialist views and theories of authority.

5.1 Hobbes’ Fool

HCA, similar to other forms of Hobbesianism, holds that one has most reason to act so as to best promote her desires; for ease of discussion, I will summarize this implication by referring to it as a ‘desire-promotion view’, or DP. Now, of course, the first worry with any such view is that it makes it implausible that most people, much of the time, have most reason to act morally. The arguments of the previous four chapters, however, were intended to show that HCA can
respond to this initial worry. While I have admitted that it is *not* likely the case that everyone, always has most reason to act morally, the connection between the conditions for intimacy and moral action (as well as more direct interest in moral living) makes it the case that at least some of us – and many of those with whom we live and engage – have most reason to act morally, much of the time. This is because, even if we do not desire to act rightly, we *do* (many of us) centrally desire to have intimate relationships, and the Scanlon-inspired argument of Chapter 1 claimed to demonstrate that moral action promotes the possibility of genuine intimacy.

Those arguments may not have convinced then, and so do not convince now. Regardless, rehashing them is not the present goal. Instead, the goal of the current chapter is to ask whether HCA, as a Hobbesian view, can succeed *even if we grant* the arguments above.

The Problem of the Fool is intended to show that the answer is ‘no’: even if we grant the contentious arguments of the previous chapters, we cannot pull stable moral reasons out of considerations of self-interest. The first move of this challenge was made by Hobbes with the famous words, “The fool hath said in his heart, there is no such thing as justice, and sometimes also with his tongue” (1968, Ch. 15). Why would the Fool say this? Not, Hobbes says, because he denies that there are covenants and that breaking a covenant may actually be called injustice; but rather, because “he questioneth whether injustice…not sometimes stand with that reason which dictateth to every man his own good” (1968, Ch. 15). Since reason ‘dictates’ to every man his own good, the question here is what we should say of those cases in which one’s ‘own good’ – or as I have been writing, one’s ‘desires’ – dictate paradigmatically immoral action (the breaking of one’s covenant). In Hobbes’ version of the problem, the Fool asserts that reason
sometimes dictates injustice, which leads the Fool to say, at least in his own heart, that there is no such thing as justice (since surely justice must be dictated by reason).

Abandoning Hobbes’ language and substituting the framework of reasons theory here utilized, we can see the same challenge arise: if I say that one has most reason to act in those ways that best satisfy her desires, then anyone who sees benefit (in terms of her desire satisfaction) by breaking the moral law is in fact commanded by reason to do so. And while HCA can respond by pointing to the ways in which moral action *tends* to be in one’s interest (after all, very many of us want genuine intimacy), my critic needs only one plausible example in which it is *in fact* in one’s interest to act unjustifiably. And this is not difficult. One need only ask *why* moral action tends to promote intimacy; and whatever I answer, the critic can come up with an example in which that feature is missing from the case. In the instance of HCA – as with many Hobbesian views – the explanation behind why one has reason to act rightly has to do with the impact on one’s relationships when others *find out* about the immoral action. And so my opponent need only one case in which an actor is virtually certain that her unjustifiable act will never be found out. Thus, if I could kill a competitor with a near guarantee that no one would ever know, don’t I have most reason to do so, given that removing the competition is likely to promote my desires, and the normal way in which unjustifiable actions are counter to my interests (that others will find out and for that reason not form relationships with me) does not apply?

Although there may be many ways to attack the Fool’s reasoning, I think that the general problem is both serious and deep. If we concede that what one has reason to do is determined at every particular choice point, then the Problem of the Fool looks to me like it may be devastating
to HCA, and likely to all forms of Hobbesianism. Fortunately for those views, however, there are good reasons for thinking that the evaluation of token acts is a bad strategy for promoting one’s desires.

5.2 The Act-Disposition Insight

The seeds of the solution to the Problem of the Fool can be found in Hobbes’ own writing, although not yet in an entirely convincing form. After laying out the Fool’s rationale, Hobbes says of it that “[t]his specious reasoning is nevertheless false” (1968, Ch. 15). The Fool’s mistake, Hobbes argues, is in thinking it reasonable to predict gain by breaking one’s covenants, as the danger is very real: covenant-breakers will not be allowed into society, says Hobbes, or if they are, it is “by the errors of other men, which he could not foresee nor reckon upon, and consequently against the reason of his preservation” (1968, Ch. 15). The idea present in Hobbes’ response here, onto which I will build throughout the chapter, is that, even though it may in fact turn out to be advantageous for a particular agent to act impermissibly, it is not reasonable for one to believe this to be the case. Instead, each agent has most reason to adopt some strategy other than attempting, at each moment, to best promote her desires.

In the hands of David Gauthier, the basic Hobbesian insight is made more precise. “A person disposed to violate his covenants cannot be admitted as a party to cooperative arrangements,” says Gauthier, “and so such a person cannot rationally expect to reap the benefits available to cooperators. Even if his particular breaches of covenant would benefit him, yet the disposition that leads him to such breaches does not” (1986, 162). The key move in responding to the Fool, Gauthier thinks, is in changing the conversation from the rationality of actions to the rationality of dispositions. Hobbes’ solution can thus be recast in the following way: although
one who commits unjust actions may occasionally be let into society, this would be the case only as a result of error, and relying on such errors is a very bad strategy for doing well in life.

Instead, one ought to stop assessing the reasons in favor of individual, just actions, and instead, respond to the reasons in favor of becoming a just person – or, one who is disposed to act justly.\(^2\) What Hobbes’ argument then needs, according to Gauthier, is the further claim that it is rational to act justly, or to ‘perform one’s covenant’, even if doing so is not in one’s interest, so long as adopting the disposition to so act was in one’s interest (1986, 162). It is this stronger claim that Gauthier makes on behalf of his own theory, and that I will now make on behalf of HCA.

At this point, I will abandon Hobbes’ language of ‘justice’, and I will avoid importing Gauthier’s decision-theoretic language of ‘constrained maximization’ and ‘rationality’ as well. Instead, I will apply the Hobbesian lessons to HCA. As previously articulated, HCA claims that many of us, much of the time, have most reason to act justifiably, because doing so promotes our

---

\(^2\) David Boonin-Vail, in his (1994), takes the argument one step further, claiming that Hobbes is in fact best understood as a kind of idiosyncratic virtue ethicist. On this view, the character traits or dispositions that conduce to one’s survival are virtues, but they are virtues as a result of their instrumental (rather than intrinsic) good (which is why the view is idiosyncratic). If Boonin-Vail’s interpretation could be sustained, it would solve the Problem of the Fool by reference to dispositions, just as Gauthier wants. However, there are two worries. The first is that Mark Murphy argues convincingly that Boonin-Vail’s interpretation cannot be sustained (1996). Second, it is unclear how helpful such a move would be for a Hobbesian, as Rosamond Rhodes points out, since the same criticism that Boonin-Vail levels against the rule-egoist version of Hobbes – that adhering to a rule when doing so is not to one’s advantage amounts to a kind of ‘rule worship’ – can be leveled against his view (1996, 867; see also Murphy 1996, 651-652). Should one act on her honest disposition when so acting is not to her advantage? It seems that a virtue theory had better say yes. But since the virtues are justified by appeal to one’s advantage, then recommending that one act against her advantage for the sake of following her disposition seems tantamount to ‘disposition worship’. Notice that, though Boonin-Vail’s concern is with a first-order moral theory rather than a theory of reasons, a version of this same problem will haunt HCA in what follows.
desires to live justifiable lives and to have intimate relationships. However, the Fool here objects that at least sometimes, it will better promote his desire to act in unjustifiable ways, and so the Humean grounding of the view seems to imply that he has most reason to act unjustifiably. The move currently under investigation, however, suggests that it is not, in fact, the token justifiable actions that make the link between justifiability and one’s desires (particularly, for intimacy) plausible. Instead, what plausibly best promotes one’s desires is to be a justifier – that is, to be disposed to acting justifiably, or as I sometimes put the point: to have a standing commitment to acting justifiably.

To motivate the general move from act-recommendation to disposition-recommendation, consider another kind of normative theory: hedonistic egoism, which holds that one ought to do whatever makes one happiest (henceforth: simply ‘hedonism’). And now let’s consider the well-known paradox of hedonism, which holds that the single-minded pursuit of happiness may render one significantly less happy than pursuing some other lifestyle. If we interpret the hedonist’s dictum to require the single-minded pursuit of happiness, then we are left with the paradoxical conclusion that being a hedonist is a good way to lead a predictably unhappy life. Why? Well, because many of the activities that make people happy – relationships, long-term projects, raising families, etc. – require many, many token actions that are hedonic sacrifices. Further, humans may turn out to be quite bad at judging when short-term hedonic sacrifices will have a long-term hedonic benefit, and so the actual attempt to calculate and promote one’s hedonic payoff may be self-undermining. If I adopt the strategy of always choosing the action that has the highest hedonic payoff, then, I will likely never have good friends, will not have a happy family, and will likely never engage in meaningful, long-term projects; in addition, I will
likely make mistakes in judging what will make me happiest. In the language I will use below, being an ‘act-hedonist’ is a very bad strategy for promoting the grounding good of hedonism; a hedonist, instead, should become a ‘disposition-hedonist’ – or, one who adopts the set of dispositions that will likely lead to the most happiness in the long run.

The connection between hedonism and DP theories should now be clear. Recall from Chapter 4 that all DP Theories can be accurately summarized as endorsing:

DPT: One has most reason to do what will best promote her desires.

But now notice that DPT is vague in exactly the way that hedonism is, as it recommends attaining some end (the best promotion of one’s desires), but leaves it open as to how one should arrange her life so as to accomplish this end. If we then take a ‘strategy’ to name some method for adhering to a normative theory’s recommendation, then DPT, like hedonism, is open to at least two strategies. For the sake of best promoting one’s desires, one might adopt

Act-DPT: Attempt, at each choice point, to act in the way that would best promote your desires;

or, alternatively, one might adopt

Disposition-DPT: Develop the dispositions that would, in the long run, best promote your desires.³

³ I want to head off one misunderstanding early: ‘act-‘ and ‘disposition-‘ seem like appropriate prefixes for the views that I will be assigning them; however, doing so may also generate misunderstanding. As Peter Railton discusses in the context of ‘indirect consequentialism’, the distinction that I am utilizing here cuts across the moral theoretic distinction between, say, ‘act’ and ‘rule’ forms of consequentialism. This is because both act- and rule-consequentialism provide criteria of rightness (and they conflict with one another); the same is not true of ‘act-hedonism’ and ‘disposition-hedonism’, as they agree on the truth conditions for right action – they just disagree on the acceptance conditions (1988, 116).
In the context of HCA, these candidate strategies are competing recommendations for how one can best promote her desires. What they assume is that there are, in fact, desires that one might promote by acting in certain ways, and desires that one might promote by developing certain dispositions. It also assumes that there are at least some cases in which one cannot do both – that is, that the adoption of one strategy precludes adopting the other. Now, since these are strategies for following the recommendations of a DPT, the desires referenced ground reasons to act or to develop dispositions; and so the question of which strategy one ought to adopt is equivalent to the question of which set of reasons is weightier – that is, to which set of reasons one should respond. For ease of discussion, we can say that Act-DPT calls on the agent to respond to token reasons, while Disposition-DPT calls on the agent to respond to strategic reasons.4

That these are genuinely competing strategies can then be seen by noting that a commitment to respond to one’s token reasons – that is, adopting the strategy of Act-DPT – would plausibly prevent one from becoming (adopting the disposition of): a friend, a good neighbor, a reliable employee, a law-abiding citizen, a loyal group-member, etc. For each of these dispositions almost certainly requires occasionally acting because it’s what that kind of

4 To be clear, then: in the same way that the Humean thinks defeated reasons remain in the context of token reasons vs. other token reasons (i.e. while I have most reason to continue writing, that doesn’t mean I no longer have any reason to take a coffee break; I absolutely have reason to take a coffee break!), whichever strategy turns out to be the best will not entail that reasons of the other kind disappear. The question is whether one might best promote her desires by doing something other than attempting to promote her token desires. If the answer is, as I will argue, ‘yes’, then one will have weightiest reason, and so most reason, to act in this other way. But – a point to which I will return – this leaves the (often very weighty) token reasons as normative residue. See further 5.7 below.
person does, despite having most token reason to do otherwise. If I never do a favor for any of my friends, simply because doing them would sub-optimally satisfy my token desires, then it starts to look like I’m not actually a friend to these people. And if I do at work precisely what satisfies my token desires, then the chances of me being a reliable employee are very slim. Indeed, most of the dispositions and roles that we take to be valuable require that we at least sometimes act in accordance with the expectations of that disposition or role, despite having most token reason to do otherwise.

HCA, then, is a DPT, and so entails that an agent has most reason to do what best promotes her desires; but it is not clear whether responding to one’s token reasons is the best way to follow that directive. Indeed, one of the central arguments of this chapter is that, like in the case of hedonism, it is plausible that such token evaluation is not the best way to achieve the best promotion of one’s desires. Instead, it is plausible that one best promotes one’s desires by (at least sometimes) responding to strategic reasons, rather than solely focusing on token reasons. And like in the case of hedonism, this move is motivated in two ways: first, there are plausibly some desires that are best promoted by being a kind of person; and second, because humans are liable to a host of errors and biases, we are likely to better succeed in promoting some of our strongest desires by attempting to follow Disposition-DPT than by attempting to follow Act-DPT. Thus, we are likely to best promote our strongest desires by adopting Disposition-DPT, and so that is what we have most reason to do.5

5 This dual method of argument will be relevant throughout this chapter, as it will allow me to solve two, difficult problems for my solution to the Fool. First, this distinction in method will reappear in 5.7 as an explanation of my solution to what I call the Disposition-Act Problem. Since the first method of argument explains a first-order reason to develop a disposition, while
More particularly, I will make the Hobbesian argument that consideration of one’s desires in fact supports the strategy M-Disposition-DPT: Develop the ‘moral disposition’ (differently defined for each theory: covenant-keeper for Hobbes, constrained maximizer for Gauthier, justifier for HCA).

This understanding of HCA (and DP Theories more broadly) will then allow me to defeat the Hobbesian’s challenge of the Fool if I can make plausible two claims: (1) that one better promotes her desires by adopting M-Disposition-DPT than by adopting some other strategy; and (2) the Fool follows some other strategy. While I will begin to make this case by suggesting that the Fool in fact follows Act-DPT, we will quickly see that the strongest challenge comes from a Fool who adopts a much more complex strategy, and so adequately responding on behalf of the Hobbesian will require investigation of several alternative strategies.

5.3 Defeating the Fool

In defense of (1), I have available both the typical Hobbesian arguments, as well as an argument unique to HCA. The typical Hobbesian arguments are basically those above, beginning with Hobbes’ original claim that those who act every time to promote their own interests will be led to break their covenants, and covenant-breakers would not be allowed into society; or that if they are, that this would be due to others’ error (on which we should not

the second explains a second-order reason to deliberate in a particular way, I argue that the dispositions justified by strategic reasons, at least in these cases, come packaged with an explanation of why one has a reason to act on the disposition that she has most reason to have. Then, in section 5.8, I suggest that, although these two methods of justification jointly support the strategy of being a justifier, it’s possible for them to come apart. And since the second argument is the one that justifies a reason to act on one’s disposition, those dispositions justified only by the first argument do not include a reason to so act.
Gauthier expands on this insight to suggest that those disposed to violate their agreements will not be allowed into beneficial, cooperative endeavors. (1) is true for HCA, then, if being a justifier is the best strategy for best satisfying one’s desires. Now, HCA did not focus on desires to be let into society or cooperative agreements, but the Hobbesian arguments above still have some relevance: it does seem plausible, for traditional Hobbesian reasons, that a single-minded attempt to satisfy one’s desires may well be self-defeating. For many of the things that we want are better furthered by a commitment to acting justifiably than by a disposition to always act so as to best promote one’s desires. And so being a justifier seems, for traditional Hobbesian reasons, plausibly to be a better strategy at least than Act-DPT.

HCA has an additional reason to think that M-Disposition-DPT is the better strategy, though: namely, that it has already identified the formation of a disposition as what best promotes the moral motivational desire. According to the complex moral motivational story told in Chapters 1 and 2, one has reason to act justifiably, because she has reason to be a justifier\(^6\) (or, because she has an independent desire to live justifiably); this reason to be a justifier, then, is what is explained by the desire for intimacy. According to the Scanlon-inspired argument concerning intimacy, some class of the closest, most enjoyable and meaningful relationships in our lives (‘genuine’ intimacies, not merely drinking buddies or casual sex partners) are reliably hindered by a lack of commitment to justifiability because, as Scanlon says, being too close to such a person would be ‘unnerving’ for the other (1998, 164). What we can now see is that such

\(^6\) It is worth flagging, however, that it is precisely this ‘because’ that Derek Parfit and Bryan Skyrms will question in 5.4 below. What I will call the ‘Disposition-Act Problem’ is what I previewed as a challenge to the complex moral motivational story in Chapter 1 (f. 7).
an argument fits well within a Hobbesian, dispositional framework: important goods (such as intimacy) are reliably blocked (if one were to gain them, it would be from the error of others) from those not disposed towards justifiable action.

The considerations above constitute the beginning of an argument in favor of (1); what it demonstrates is that being a justifier is plausibly a good strategy, and likely better than some. What would complete the argument for showing that it’s the best strategy is a comprehensive evaluation of the competition. While comprehensiveness may be a bit much to ask for, I will here attempt to complete the argument for (1) as best I can by evaluating (2) – that is, by comparing the strategy of being a justifier to the candidate strategies that the Fool might be thought to adopt. Although I will conclude that I cannot show decisively that (1) is true, I can both show that (2) is true, and that (1) is a reasonable conclusion. It is thus reasonable to conclude that HCA can answer the Fool.

At first, it looks as though (2) ought to come easily: after all, isn’t the Fool simply one who follows the advice of Act-DPT? Perhaps it initially seems so, but in response to the argument above, one may wonder about a more sophisticated Fool. The advocate of M-Disposition-DPT has argued above that it is superior to Act-DPT as a strategy for pursuing one’s interest. And maybe, my opponent may concede, this is true. But those surely aren’t the only options. It seems that the disadvantages associated with Act-DPT stem from others’ knowledge of one’s disposition. But doesn’t that imply that the best strategy would actually be neither of those considered, but instead, one with which the agent gained the benefits both from appearing
to be a justifier, and with occasionally not acting as a justifier?\footnote{Recall that this was the challenge that Glaucon posed to Socrates in Plato’s Republic (2004) as well. For those interested to ground moral reasons in self-interest, the challenge of this sophisticated Fool has been around for quite some time.} According to this challenge, a stronger contender to Disposition-DPT is actually:

Fool-DPT: Act as though you have adopted M-Disposition-DPT whenever failing to do so would be discovered; in all other cases, act so as to best promote your token desires.

This is the real challenge of the Fool. It does appear that the benefits of M-Disposition-DPT are due to the appearance of being committed to justifiable action; and further, acting justifiably will, in fact, sometimes come with immediate costs to the satisfaction of our desires. It appears, then, that if Fool-DPT is a real candidate strategy – that is, if it is possible to appear sufficiently like a justifier so as to reap the benefits while not actually having the disposition – then Fool-DPT would be the best strategy.\footnote{This seems like a good place to emphasize that the question of which strategy is best arises only because of our lack of certain abilities. That is: the point that the Fool correctly raises here is that, if one was able to perfectly respond to all of the weightiest reasons – both strategic and token – then that is clearly what he ought to do. Thus, the question of strategy is interesting only because it’s not clear that the ‘best’ strategy is possible; that is, it is not obvious that one can gain the benefit of promoting the grounding desires for strategic reasons while promoting the desires grounding token reasons. Now, as should become obvious, the impossibility of responding to some token reasons while promoting some (much weightier) strategic reasons does not imply that the token reasons disappear. Thus, it may be the case that one has most reason to adopt a disposition that recommends acting against what one has most token reason to do in some instance. All of this will be discussed more, below, but it seems helpful to flag here.} However, I do not think that Fool-DPT is a real candidate, and explaining why will require that I borrow one more tool from Gauthier.

It would be easy to explain why Fool-DPT is impossible for humans to follow if we were, as Gauthier puts it, transparent – that is, if our dispositions and intentions were clear for everyone to see (1986, 173-174). If we were, then we would be able to read off the faces of the
Fools of the world that they were not in fact justifiers, and that they will only keep their word if
doing so is beneficial to them, and so those interacting with Fools would not confer upon them
the advantages of being a justifier. In such a world, following Fool-DPT would just be to follow
Act-DPT, as any Fool would not be offered the opportunities for mutually advantageous
cooperation.

Humans are not, alas, transparent. At least some people can lie convincingly, at least
some of the time, and so it is at least possible that someone could follow Fool-DPT. However, I
agree with Gauthier that people tend to be translucent, which is to say that we are not
transparent, but we are not opaque either. A translucent individual’s dispositions and intentions
regarding cooperation and interaction may be ascertained by others, “not with certainty, but as
more than mere guesswork” (Gauthier 1986, 174). And this, I think, is enough to determine that
we (at least the vast majority of us) ought to adopt M-Disposition-DPT over Fool-DPT as our
strategy for pursuing our own interests. Why? Because the stakes are high, and translucency
introduces too much risk. Disposition-DPT offers an extremely promising way to promote
important, central desires, with a very high degree of probability. Fool-DPT, on the other hand,
could be disastrous, as the cost of misjudging our own translucency could be a lifetime without
friends, family, or even one’s freedom.

To use a well-known example, consider Parfit’s ‘Desert Break-Down Case’.9 Suppose
that my car has broken down in the desert, and I desperately need a ride home to survive.
Unbelievably, on this virtually deserted highway, another motorist comes along. After flagging

---

9 The example is found in Parfit (1984, 7), but I take the name from Gauthier, in his (1997, 26).
him down, I explain my situation and offer a large reward should he give me a ride home. Of course, my money is at the house, so I will have to give it to him after we arrive there. If I am translucent (which believe me, I am), then failing to have a disposition towards keeping my word, even when it’s to my disadvantage, would be disastrous in this situation. If I know about myself that, once safely home, I will have to calculate whether it’s to my advantage to pay a large reward to the motorist, then there is a real danger that I will fail to hide this fact from the driver. In this scenario (and every other scenario with a similar pay-off structure), failing to have a trustworthy disposition would be very risky. Acting as though one has a disposition is not the same as actually having the disposition, unless people are opaque, which they are not.

Now it may be the case that there are a few individuals who are genuinely opaque – perhaps the very best con artists fall into this category. The next question, then, is whether a person is a good enough judge of her own opacity to justify adopting the Fool’s strategy. If we consider a realistic psychology – incorporating any biases, irrationalities and errors that one is likely to have – do we believe that most of us would be competent judges of our own opacity or translucency to others? Especially if we now consider the risks considered above: would it be rational to count on fooling others in order to secure some of life’s most important goods? As is likely obvious, the question is largely an empirical one, and I have no idea how to settle it decisively. But my suspicion is that such a move would not be rational. There are very few gifted con artists, and those rare individuals who are ‘naturals’ at deception do not have sufficient good evidence to ground a confidence in their own ability. It is more reasonable to believe that attempting to con others will eventually have a high price, and so one has more reason to adopt a conservative strategy in promoting her own desires.
As a final defensive maneuver, my opponent may ask why the Fool must see translucency as a risk – that is, why the sophisticated Fool cannot genuinely adopt the recommended dispositions (in the case of HCA, that of being a justifier) most of the time, but abandon it when doing so would be useful. Perhaps, one might argue, the Desert Break-Down Case is convincing because defecting against the helpful motorist seems to require such a malicious strategy; that is, the case convinces because one would seemingly defect only if she had a commitment to calculating her self-interest payoffs when interacting with others. And that kind of scheming disposition seems like exactly the kind of character trait that may be difficult to hide (for translucent beings like us). But not all unjustifiable acts seem to require such a commitment. So we can imagine instead the Dropped Wallet Case, in which a woman finds a dropped wallet in a public bathroom. With the door locked, she quickly looks inside to see over $300 in cash. She knows that there are no cameras around and no record that she was there. There appears to be an exceptionally high probability that she could simply take the cash, leave the wallet, and walk away without anyone ever knowing, and she would be $300 richer.

This final challenge is the most difficult, as it concedes Disposition-DPT, but challenges M-Disposition-DPT. Perhaps, the objection goes, one ought to develop settled dispositions as a strategy for promoting one’s desires; and perhaps, it’s even the case that the disposition should be largely towards justifiability; however, sometimes scheming isn’t required, and an agent will be presented with an unjustifiable act that has a very high benefit-risk ratio. In this case, the Fool claims, an agent who endorses M-Disposition-DPT is very likely to do worse than an agent who endorses
F-Disposition-DPT: Adopt the ‘moral disposition’; however, if presented with an opportunity that you cannot help but recognize as overwhelmingly likely to be beneficial, take it, regardless of whether doing so is what one with the moral disposition would do. The adherent of F-Disposition-DPT would take the $300 from the wallet, and likely be $300 richer without incurring any costs. The adherent of M-Disposition-DPT would (under plausible assumptions about morality’s requirements) either leave the wallet, or attempt to find out to whom it belongs, so that it could be returned. The Fool’s most difficult challenge is that it is very difficult to see how, given the existence of such cases, consideration of one’s desires could recommend M-Disposition-DPT over F-Disposition-DPT.

At this point, I think all the Hobbesian can do is to admit a kind of conservatism, or risk-aversion. It doesn’t seem an unreasonable conservatism to me, but the basic suggestion is straightforwardly risk-averse. To see how, let us now assemble our candidate strategies for pursuing one’s interests. The Hobbesian’s initial move was to show that M-Disposition-DPT is superior to Act-DPT, but it has become clear that this is not sufficient for concluding that one ought to adopt the former. Instead, these two strategies represent poles of a spectrum: both are ‘pure’ strategies, recommending either a pure commitment to token evaluation, or a pure dedication to the moral disposition. But between these opposing, pure strategies, is a full range of ‘mixed’ strategies, incorporating both act- and disposition- elements. The most sophisticated form of the Fool’s objection, then, points to the graphic below and suggests that one of the mixed strategies on the left (M₁ or M₂) seem more likely to be the best overall strategy.

\[
\text{M-Disposition-DPT} \rightarrow (M_1) \rightarrow (M_2) \rightarrow (M_3) \rightarrow (M_4) \rightarrow \text{Act-DPT}
\]
I do not believe that it can be shown decisively that one of these strategies is the best overall. The final suggestion of the Hobbesian is that a reasonable risk-aversion combines with the ease of implementing a pure strategy (over a mixed one) to count in favor of the pure strategy. In short, given a person’s fallibility, it will be very difficult to determine with certainty whether Dropped Wallet is a case that favors M₁ over M-Disposition-DPT. Further, the Hobbesian worries that, should one take the money, that success (should one find it) will boost one’s confidence in benefitting from calculating self-interest, and will generally erode the moral disposition, leading to riskier behavior. In contrast, the adoption of the pure disposition is easy to implement, self-reinforcing, and brings with it the kind of conservative assurance of high payoff with low risk. Of course, this is all in addition to the kind of side-benefits that any Hobbesian will point to, such as the satisfaction of returning one’s possessions and the boost in reputation that comes with a dedication to justifiable living. In short: a clever person may in fact gain a little by sliding ever so slightly away from the left pole of the spectrum and into a mixed strategy; but reliably predicting that gain is difficult and dangerous, whereas simply adhering to the pure strategy of M-Disposition-DPT is easy to implement, self-reinforcing, and it carries many side-benefits as well. These considerations are not decisive in favor of M-Disposition-DPT, but I think that they render reasonable the conclusion that being a justifier is the overall best strategy. At the very least, the arguments here canvassed seem to suggest that it is the best strategy for any agent with the common characteristics explored above – that is, one who is
translucent, risk-averse, and who lacks confidence in her own ability to judge when a case is likely to be one of those that would allow a mixed strategy to reliably benefit.\(^{10}\)

This concludes my Hobbesian argument for the claim that considerations of desire-promotion recommend the strategy M-Disposition-DPT, and so for the view that one has most reason to adopt the moral disposition as a means to best promoting her desires. As a result, I now hold that HCA is the view not that those who centrally desire intimacy (or to live justifiable lives) have most reason to act justifiably; but rather, it is the view that those same individuals have most reason to be justifiers, or to develop a standing disposition or commitment to living a justifiable life.

5.4 The Disposition-Act Problem

\(^{10}\) It has been suggested to me that, if my argument justifies only these weaker conclusions, we might still make good sense of important moral phenomena. For instance, if psychologies are such that some agents may in fact benefit from adopting M\(_1\), then we might suppose that some agents will have most reason to take the $300 in Dropped Wallet. However, it seems very unlikely, even to my opponents, that M\(_1\) or M\(_2\) would ever recommend murder or kidney theft, due to the very high costs and the serious risk. The suggestion, then, is that my analysis of strategies here explains why we take murder and kidney theft to be much worse than keeping the money in Dropped Wallet, but also why some may not even share the intuition that you are obligated to leave the money in Dropped Wallet. Perhaps, in this case, tracking down the owner would be supererogatory, but that one might, in fact, have most reason to take the money. Although I’m not sure that I want to take on this explanation fully, it is worth continuing to think about. Thanks to Jake Earl for the suggestion.

Relatively, Steve Kuhn, in his (1996), uses a version of the Dropped Wallet Case to illustrate the way in which we might need a doubly ‘indirect’ moral theory (much more on indirectness below). Perhaps, he suggests, the set of moral rules that we should adopt would actually be complicated in the following sense: perhaps we would all be best off if everyone were to keep any money found less than $20, but to turn in any larger amounts found. However, in order to accomplish the adoption of these rules, perhaps we need people to feel guilty keeping anything over $10; and in order to inculcate that level of guilt, perhaps we need to teach our children that they ought not to keep even $1. Although I will not adopt Kuhn’s particular form of indirectness (that the correct moral rules are the ones that we ought to teach others), it should become clear below that I take this general move to indirectness to be well-motivated.
It is unlikely that the arguments above will convince the skeptic. If one is generally skeptical of a Hobbesian approach, I have offered little to change one’s mind. However, I do think that the Hobbesian should embrace something like the argument above, and therefore accept the position that considerations of self-interest generate reasons to adopt a disposition towards paradigmatically moral action – whether this is a classically Hobbesian disposition towards ‘covenant-keeping’, Gauthier’s ‘constrained maximization’, or my proffered ‘being a justifier’. And if this position is accepted, then I think the Fool has been answered.

Suppose, then (optimistically), that my argument above does in fact justify the strong conclusion – that one has most reason to adopt M-Disposition-DPT, even over similar positions like M1. The Hobbesian’s next challenge is that the solution to the problem of the Fool has generated a new problem – and this one may be even more difficult to solve. In short, the new challenge is that, having successfully shown that one has most reason, not to act, but to develop dispositions, we may have generated a problematic normative gap. The basic observation here is that a reason to be a justifier is not a reason to act justifiably. Those reasons seem related, sure, but they are not identical. And so the question arises: having set out to explain moral reasons – that is, the reasons one has to act rightly – have we just now argued ourselves out of any such explanation?

Let us make explicit the problem. The arguments above purport to justify the claim

A. Because one has most reason to best satisfy her desires, one has most reason to be a justifier.

However, the desideratum of explaining moral reasons by reference to self-interest would be satisfied only if we justified
B. Because one has most reason to best satisfy her desires, one has most reason to act as a justifier.

It may seem plausible that B simply follows automatically, perhaps even conceptually, from A. Gauthier himself seemed initially to think this, moving to his version of B with the comment, “[i]f [an actor’s] dispositions to choose are rational, then surely her choices are also rational” (1986, 186). After all, could one be said to have the moral disposition if she didn’t act morally?

Unfortunately, B does not come so easily. There are at least two kinds of considerations that suggest we need an argument for the move from A to B. First, one act does not make a disposition, and so presumably one act does not break one either. If one has most reason not to be a liar, then one satisfies that requirement by not being a liar; but does lying once make one a liar? It is not obvious that it does. As Christine Korsgaard puts the point in her discussion of practical identities: one can occasionally act against who she is and still ‘find her way home’ (1996, 102). But if one can avoid being a liar even if she lies once, then it does not seem that having most reason not to be a liar entails having most reason not to lie in some particular, token instance. After all, I may do what I have most reason to do (avoid being a liar) and yet lie at some point.

Secondly, there are two kinds of cases, made famous by Derek Parfit and Bryan Skyrms, which purport to show that no middle premise will allow the derivation of B from A for all cases.\(^\text{11}\) The first is the case of *rational irrationality*, in which the fact that one rationally ought

\(^{11}\) Note, however, that since this is the task of the examples – to show that are *some* cases in which B cannot be derived from A – they block only those arguments that claim that B *always* follows from A. Gauthier’s conceptual claim, then, is challenged by these examples, as would be any substantive claim of the form “one ought to act on the dispositions one ought to have.”
to adopt some disposition does not seem to imply that acting on the disposition is rational.

Consider: 

_Schelling’s Answer to Armed Robbery:_ A robber threatens Schelling, saying that unless he unlocks his safe, the robber will start to kill his children. It would be irrational for Schelling to ignore this robber’s threat. But, even if he gave in to the threat, there is a risk that the robber will kill the whole family, to reduce his chance of being caught.

However, Schelling does have an irrationality serum, which will make him temporarily act in completely unreasonable ways, and the robber will be able to see that Schelling is unable to respond to reason. Schelling, believing correctly that doing so is his best available strategy, drinks the irrationality serum in front of the robber (Parfit 1984, 12-13; 2001, 84).

According to Parfit, Schelling rationally ought to drink the serum, and so adopt and have the disposition of being irrational. He should do this, even knowing that as a result, he will do terrible things, like beating his children because he loves them, and burning his manuscripts because he wants to preserve them (2001, 86). Parfit’s claim, then, is that the rationality of drinking the serum simply does not imply the rationality of acting on the irrational disposition: beating his children and burning his manuscripts are irrational, despite the fact that it is rational to have the disposition. But the case and all of the claims do not depend on utilizing the

---

However, a more modest strategy that claims that one *sometimes* ought to act on the dispositions one ought to have would not necessarily be troubled by these cases; rather, such a view would simply owe an explanation of why Parfit’s and Skyrms’ cases are not among those in which the reasons to adopt a disposition generate reasons to act on it. In section 5.8, I will suggest that my own solution to the DA problem follows this more modest strategy, and so I provide an explanation of Parfit’s and Skyrms’ cases.
language of rationality – we could translate it all into the framework of what Schelling has most reason to do. And Parfit’s claim will still be that, given the details of the case, Schelling has most reason to drink the irrationality serum, but that neither Schelling nor anyone else could have most reason to act in the ways he then acts. If Parfit is right, then B simply does not follow from A. It is sometimes the case that one has most reason to adopt a disposition that will result in one’s acting in ways that one has no reason to act, and most reason not to act.

Perhaps even more difficult are cases of threat enforcement. As both Skyrms (1996) and Parfit (2001) argue, the structure of certain scenarios can make it such that developing the disposition of being a threat-enforcer is what one has most reason to do. We can imagine this to be the case in international politics, in order to control rogue nations (Skyrms 1996), or even just among individuals, in order to keep from being taken advantage of (Parfit 2001). In either case, in an effort to prevent disaster, one finds that she must: threaten to do something disastrous (say, launch one’s nuclear arsenal) if the other party performs some act; and convince the other party that the threat is genuine. However, if we are translucent, as Gauthier and I have assumed, then accomplishing the latter will require adopting the disposition to actually follow through with one’s threat – that is, to become a threat-enforcer. The challenge comes when the other party, perhaps out of sheer insanity, performs the forbidden act. Whatever it was, the damage is now done, and the threat-enforcer has the disposition to launch one’s nuclear arsenal. But if the cost of doing so is global annihilation, then clearly one has most reason not to launch. It appears, then, that we have another case in which one has most reason to form a disposition (that of being a threat-enforcer) that one has no reason to act on.
The gap between A and B results in what I will call the Disposition-Act Problem, or DA Problem. In response, HCA – or any Hobbesian view that takes the move to dispositions in order to respond to the Fool – must provide an explanation for how reasons in favor of adopting a disposition can generate reasons in favor of acting on that disposition. In the following sections, I will outline a series of arguments that I believe allow the Hobbesian successfully to make this move.

5.5 The Class of Indirect Normative Views

At this point I want to define a class of normative theories, held together by an interesting structural similarity. As will become obvious, I think this is worth doing, not only because it is an interesting and perhaps fruitful note on similarity across otherwise disparate views, but because I believe that the shared structure has invited a shared problem. Each of the views I will here discuss faces, I will argue, its own version of Fool-DA Problems, even if they haven’t been recognized as such.

The class in question is that of ‘Indirect Normative Theories’, or INTs. Some theory is an INT if it has the following structural property: it takes some normative claim, C, to be true, but it simultaneously holds that, in order to best follow C, one ought to adopt some strategy other than following the recommendations of C at each choice point.\(^\text{12}\) In order to focus on just the

\(^{12}\) Terminology in this discussion is difficult, as most of the natural terms have been used for related, but often subtly different positions. So, similar to how I adopted ‘act’ and ‘disposition’ language, while acknowledging that it cuts across some of the classic moral theory discussion, here too, I want to adopt the language of indirect theory, but note some important relationships. Richard Brandt, in his (1988), discusses the class of ‘indirect optimific theories’. Brandt’s account of indirectness then influenced the view of Steven Kuhn (1996), who articulates an indirect contractarianism. Given that I will go on to discuss the way in which my own contract view is indirect, it might be expected that I simply adopt something like Kuhn’s view. However,
the kind of indirect theory that Brandt and Kuhn discuss is importantly different from the class of views that I am delineating. According to Brandt, an indirect optimific view (roughly) holds that an act is permissible if it would be best for everyone to permit it, and an act is obligatory if it would be best for everyone to require it (1988, 342). Kuhn, then, argues for an indirect contractarianism, rather than an indirect optimific view, on which we morally ought to follow the rules that rational people would agree to teach to themselves and to others (1996).

There is a clear relationship between Brandt’s and Kuhn’s views, and the views that I am here calling indirect. The relationship is due to the fact that all views being discussed are two-level theories: there is some ultimate justification, and then a question about how best to achieve the justified result. The difference, however, is how the levels are specified. Compare Brandt’s indirect optimific view with another optimific view that is indirect under my description – that of Peter Railton’s consequentialism. On Brandt’s view, the indirectness is built into the criterion of permissibility – what is permissible is what it would be best for everyone to permit. On Railton’s view, however, the criterion of right action is direct: one is required to maximize the (pluralistic) good. The indirectness, then, concerns what strategy one takes to maximize the good (1988). One could either attempt to maximize the good through acting, at every choice point, with the goal of maximizing the good; or one could adopt a set of rules, or a disposition, with the goal of maximizing the good. While the difference is subtle, it is also important. On Brandt’s version of indirectness, once we know what it is best to permit, we know what one ought to do. However, on Railton’s version of indirectness – which follows the structure if INTs as I have laid it out – it is possible for morality to recommend a strategy (following a rule or adopting a disposition) that would lead to one violating her criterion of right action (acting at some choice point in a way that does not maximize good consequences). Thus, it is important to keep in mind the structural property I am pointing to with the language of indirectness, as it is not identical to the structural property identified by Brandt with the language of indirectness.
The second case is Utilitarian Kantianism.\textsuperscript{13} Suppose it were the case that a card-carrying utilitarian were approached by an evil demon, who vowed to exact global devastation if the utilitarian did not become a card-carrying Kantian. In this case, utilitarianism would advise that the agent do anything in his power to become a Kantian. For the sake of adhering to ‘C’ – the requirement that one produce the greatest balance of happiness over unhappiness – the utilitarian would be required to live his life by reference to the Categorical Imperative.\textsuperscript{14}

Hopefully these toy cases have made clear the structure that makes a normative theory indirect, and so it should be no surprise that HCA is an INT. At the ground level of explanation, HCA employs a proportionalist, Humean theory that makes it a desire-promotion theory; that is, according to HCA, one has most reason to do what best promotes her own desires. However, the arguments of this chapter claim to show that HCA recommends a strategy for promoting one’s own desires that requires one to do something other than attempting to promote one’s token desires – namely, it recommends the strategy of being a justifier. Thus, HCA holds that some normative claim is true – that one has most reason to promote her desires – and simultaneously

\begin{itemize}
\item \textsuperscript{13} I take this example (although not the name) from Railton (1988, 116).
\item \textsuperscript{14} Depending on the details, these cases may have the further property of being what Parfit (following Sidgwick) calls ‘self-effacing’: that is, they may be theories that provide recommendations that can be followed only if they are not known (1984, 40–42). So if the very knowledge that hedonism is true undermines my pursuit of happiness, or if the very knowledge that utilitarianism is my grounds for Kantianism undermines my dedication to the Kantian program, then both theories will recommend that one not know the true, ultimate justification. However, it is useful to see that this is a separate distinction from that of being ‘Indirect’, as it is unclear whether the theories at issue in this chapter must actually be self-effacing, or if they are merely indirect. Indeed, Peter Railton says explicitly that a ‘sophisticated hedonist’ – that is, a hedonist who adopts some strategy other than the direct pursuit of happiness – need not hide the hedonistic underpinning for his strategy from himself. On his view, if one dives into relationships, say, as a strategy for becoming happier, then “the external goal of happiness reinforces the internal goals of his relationships” (1988, 105).
\end{itemize}
holds that one ought, for the sake of promoting her desires, to adopt some strategy other than acting, at each choice-point, so as to promote her desires.

There are two other INTs that I will mention here, as it is from these particular theories that I will draw in crafting my solution to the DA Problem. The first is the indirect consequentialism defended by Peter Railton (1988). On Railton’s view, the key to defending a plausible consequentialist theory is in making clear the distinction between ‘objective’ and ‘subjective’ consequentialist theories. According to objective consequentialism, “the criterion of the rightness of an act or course of action is whether it in fact would most promote the good of those acts available to the agent”; subjective consequentialism, by contrast, “is the view that whenever one faces a choice of actions, one should attempt to determine which act of those available would most promote the good, and should then try to act accordingly” (1988, 113). Subjective consequentialism thus focuses on the actual decision procedure of an agent, while objective consequentialism prescribes no decision procedure at all – only a criterion of rightness. Railton then reserves the term objectively consequentialist act (or life) to describe those acts (or that life) that would bring about the best outcome (1988, 113).

Finally, Railton defines a sophisticated consequentialist as one whose goal is “leading an objectively consequentialist life, but who need not set special stock in any particular form of decision making and therefore does not necessarily seek to lead a subjectively consequentialist life” (1988, 114). Railton’s own view, then, is that agents ought to be sophisticated consequentialists. Thus, his consequentialist account is an INT: on what we can call his Indirect Consequentialism, one has most reason to live an objectively consequentialist life, but this does not entail reasoning in a subjectively consequentialist way. In fact, through a series of powerful
examples, Railton argues convincingly that, analogous to the hedonism case, being a subjective consequentialist would be a terrible strategy for pursuing an objectively consequentialist life. In addition to rather trivial reasons not to be a subjective consequentialist (such as the cost of calculation, or the predictability of regular outcomes), Railton is specifically interested in those cases where being a dedicated subjective consequentialist would prevent one from investing in relationships and projects that may require token sacrifices but which ultimately constitute weighty goods (1988, 114-116). Similar to the case of HCA, then, a sophisticated consequentialist’s commitment to leading a consequentialist life recommends some strategy other than acting, at all times, so as to bring about the best consequences.

The other INT that I will here discuss is Joseph Raz’s account of political authority. Although different from HCA and Indirect Consequentialism in not being a moral theory, Raz’s account shares the same structural property articulated above, which makes it susceptible to a version of the Fool-DA Problem (and which, crucially, suggests a partial solution to solving the problem for indirect theories).

The problem of political authority is that “it requires one to let authoritative directives pre-empt one’s own judgment. One should comply with them whether or not one agrees with them” (Raz 1990a, 5). If a genuine authority gives me a command, then whether or not I had any reason to do the commanded act prior to the command, I do have such a reason now – and in fact, it seems that I have overriding reason, or most reason to so act. So if my Drill Sergeant commands that I ‘drop and give her 50’, I now have a brand new, overriding reason to stop what I’m doing and do 50 push-ups. Accounting for this authority is a ‘problem’, in that it is mysterious where this new reason comes from: while it seems that such a reason must be at least
partially explained by its relation to other, independent reasons (i.e. my reason to obey my
government depends on the good reasons for employing a single, coordinating power), a
completely derivative reason would not justify *genuine* authority. If the only reason I had to
obey my government was that it was, say, the most efficient means of performing important
coordinating tasks, then should some command be inefficient in this regard, it would not be
reason-giving. Genuine authority requires that the authoritative commands are (overridingly)
reason-giving, even if some token commands do not serve to promote the grounds that justified
the authority in the first place.

Although I certainly cannot take the time fully to detail Raz’s views on political authority
here, identifying the key claim of his account will suffice for present purposes. According to
Raz,

The normal and primary way to establish that a person should be acknowledged to have
authority over another person involves showing that the alleged subject is likely better to
comply with reasons which apply to him (other than the alleged authoritative directives)
if he accepts the directives of the alleged authority as authoritatively binding and tries to
follow them, rather than by trying to follow the reasons which apply to him directly
(1986, 53).

This is the Normal Justification Thesis, or NJT. Its virtues as a theory of authority are that it
both makes it clear how the reason generated by an authority’s command depends on an
independent good (adherence to one’s independent reasons), while still allowing the ‘space’ for
genuine authority to emerge: since one is only ‘likely’ better to comply with the reasons that
apply to him, it is possible for the authority to get it wrong, and so following a genuine
authority’s interest does, in fact, make a difference to what one has most reason to do.\textsuperscript{15}

To see how the NJT is an INT, we must discover the ‘grounding’ normative claim at
issue in the context of authority. The key is in the closing phrase of the block quote above:
“rather than by trying to follow the reasons which apply to him directly.” The assumed
normative claim here is difficult to spot because it is trivial; I will formulate it as the claim that
one has most reason to act in accordance with her weightiest reasons. The very existence of
reasons, we can now see, makes possible an indirect structure, not for a ‘moral theory’, but for a
general normative theory. Analogous to Railton’s definitions in the consequentialist case above,
we can define \textit{objective} normativity as the view that ‘the criterion of what one has most reason to
do is whether it would in fact best satisfy the reasons the agent has for acting’; \textit{subjective}
normativity, then, ‘is the view that whenever one faces a choice of actions, one should determine
which act of those available would best satisfy the reasons one has for acting, and then act
accordingly’. A ‘sophisticated agent’, then, is one who’s goal is ‘leading an objectively reasons-
responsive life, but who need not set special stock in any particular form of decision making and
therefore does not necessarily seek to lead a subjectively reasons-responsive life’.

The NJT, then, establishes genuine authority if one ought to be a sophisticated agent in
the particular way described – that is, if one is better off in terms of living an objectively
reasons-responsive life by adopting the strategy of taking some agents’ (or institutions’) decrees

\textsuperscript{15} This is Raz’s ‘Difference Thesis’: although reasons generated by an authority’s commands
ought to depend in some way on the reasons one has independently of the authority’s commands,
there should not be perfect overlap. An authority is only genuine if its dictates at least \textit{may} make
a difference (1990a, 126).
as reason-giving and then responding to these reasons than she is by evaluating and responding to her own reasons independently. And the appeal of Raz’s solution is that this seems plausible in at least some circumstances. Similar to the way in which doggedly pursuing either happiness or consequentialist outcomes would be bad strategies for being happy or making the world better, trying to act, at every choice point, on the balance of reasons seems like an exceptionally bad way to live a life that best satisfies one’s overall balance of reasons. And again, some of the explanations for this are mundane – the cost of calculation, predictability of outcomes, etc. – and some are much more interesting. For instance, humans might be bad reasons-responders in very predictable ways, such that if left to our own calculations, we would reliably and predictably act contrary to our own reasons.

Consider the example of motorcycle helmet laws: it seems to me that I always have most reason to wear my helmet. However, summers can get very hot, and sometimes I’m only going down the road, and so it is not unreasonable that there might be particular token instances in which the comfort reasons combine with the safety reasons to favor not wearing a helmet. For instance: the risk of an accident on the neighborhood roads of SC might be so low, and the discomfort of wearing a helmet in August in SC so high, that for some particular set of rides, going without a helmet is what is favored by the reasons. However, because SC summers are so hot and so long, and the difference between neighborhood roads, secondary roads, and minor urban roads so difficult to quantify, I may well not be very good at correctly identifying the (few, let us suppose) instances in which reason actually favors riding without a helmet. I may regularly overweight the comfort reasons, underweight the safety reasons, as well as making generally human mistakes such as overestimating my abilities and discounting the danger. It
seems reasonable to me in the case of helmet laws, then, that a comparison between simply obeying the law and attempting to judge what I have reason to do on a case-by-case basis will reveal that I do better at responding to my own reasons by simply obeying the law. Whereas in the former case, I will very often be less comfortable than I could have been (and will sometimes act in a way not favored by the token reasons), I will wear a helmet every time the reasons favor doing so, and I will be protected any time I crash. I thus ought to take the helmet law as overriding in my deliberation, rather than attempt to evaluate and respond to the independent reasons that apply, and so the NJT would justify the authority of helmet laws.16

This concludes my foray into the class of Indirect Normative Theories. In addition to the rather uninteresting ones like hedonistic egoism and Utilitarian Kantianism, I have argued that this class includes Hobbesian views such as HCA, Railton’s Indirect Consequentialism, and Raz’s account of a Normally Justified Authority. All of these theories – different in many respects – are held together by the important structural property of taking some normative claim, C, to be true, but in simultaneously holding that, in order to best follow C, one ought to adopt some strategy other than attempting, at each choice point, to follow C. In the final sections, I will show how this structure invites versions of the Fool-DA Problem, and how borrowing from each theory’s attempt to solve these problems can aid in constructing a more complete solution for all such theories.

5.6 The Generalized Fool-DA Problem

---

16 Note from above that the NJT actually only refers to persons, and so I don’t know whether Raz himself would approve of my use of the NJT to justify the authority of a particular law. The reasoning, however, seems identical, and so the example seems compelling to me.
The Fool-DA Problem is the combined challenge raised against HCA to both respond to the Fool, and to then find a way to make sense of paradigmatic reasons for moral action (given that HCA’s response to the Fool involved the move to claims about dispositions). In what follows, I will show how the structural similarity among INTs makes a move away from the evaluation of acts attractive in each case, and so renders each such view susceptible to versions of the Fool-DA Problem.

In the case of Indirect Consequentialism, it is easy to see how a Fool might make his case. The criterion of rightness is, after all, the promotion of goodness, and so it is easy to imagine a Fool who says in his heart (and sometimes also with his tongue) that there is no justice — and perhaps no enduring projects or relationships either. The Fool would not deny that we call certain acts just, or project-based, or caring, but he would argue that rightness, which dictates to each person that he promote the good, may not always recommend just, project-based or caring actions. And in what sense does one have a project, if it must be abandoned whenever the consequences so dictate? In what sense is one in a relationship?

Given my Hobbesian argument above, the obvious response here is to move away from the evaluation of token acts and instead to focus on dispositions. And this is exactly what Railton does. He writes that the sophisticated consequentialist would recommend “cultivating dispositions that will sometimes lead him to violate his own criterion of right action,” as he should realize that certain goods are reliably attainable – or attainable at all – only if people have well-developed characters; that the human psyche is capable of only so much self-regulation and refinement; and that human perception and reasoning are liable to a host of biases and errors. Therefore, individuals may be more likely to act rightly if they
possess certain enduring motivational patterns, character traits, or prima-facie commitments to rules in addition to whatever commitment they have to act for the best (1990a, 119-120).

Of course, if the DA Problem is a threat to HCA, then it is a threat here as well: if one has, on indirect consequentialist grounds, most reason to become a loyal friend, how do we explain the reason to act as a loyal friend, especially in all of those cases where one might very slightly promote goodness in the world by not acting as a loyal friend? I believe that Railton has a partial answer to this question, to which I will return below.

In the context of political authority, the role of the Fool is played, I think, by the anarchist. The challenge of philosophical anarchism is due to Robert Paul Wolff, who argues that genuine authority is incompatible with a duty to act autonomously (1970). While the particular language of duty and autonomy may not be necessary in the current context, Wolff’s general worry is an understandable one: shouldn’t we, the anarchist asks, respond to our own reasons? And yet taking another agent or institution to be authoritative involves giving up precisely that activity. And so the anarchist Fool says in his heart that there is no such thing as authority; of course, we might call some agent an authority, but what one ought to do is dictated by reasons, and these reasons will not always recommend obedience. And so in what sense is there actual authority, if one need not obey it whenever the grounding reasons run counter to it?

In line with the pattern that ought to be familiar by now, Raz responds to the anarchist Fool by moving away from the evaluation of token acts. The best strategy for acting according to one’s reasons, given a normally justified authority, is to cease token evaluation of one’s reasons and to simply accept the authority’s commands as binding. In making this move away
from the evaluation of token acts, Raz does not, as far as I know, utilize the language of disposition. We could do so, on his behalf, and it would make good sense of the INT pattern: we could say, for instance, that in the presence of a normally justified authority, one ought to be a ‘subject’ – that is, one ought to dispose oneself to habitually obey the authority’s commands. And in fact, if the arguments given by me, Gauthier, and by Railton, are convincing, then we should, in fact, say this on Raz’s behalf.

However, if Raz were to have made this move, then he, too, would have faced the DA Problem, as a reason to be a subject is not a reason to act as a subject. And the very point of the NJT is to show that an authority can be legitimate, which is just to say a purported authority’s commands could in fact be reason-giving. But if he has shown that, then he does not face a DA Problem; if an authority’s commands are reason-giving, then a subject has reason to act on each command.

My suggestion, then, is the following: when faced with the challenge of utilizing an INT, Raz did something different from the other two theories, by focusing elsewhere than on the move to dispositions; if, upon inspection, his argument is convincing, then we may have the beginning of a solution to the DA Problem. By then incorporating some further lessons from Railton’s discussion, we may have more than a beginning – we may have the recipe for responding to Fool-DA Problems on behalf of all INT theories.

5.7 Solving the Fool-DA Problem for HCA

The difference between Raz’ account of authority and the other INTs here investigated is, most simply, that an account of authority paradigmatically involves more than one agent. While the general structure of the theories is the same, there is thus an important difference: Raz is not
(necessarily?) looking to explain the strategy of an agent, but rather the normative power of the authority that becomes part of the agent’s strategy. Although the vindication (if he is successful) of legitimate authority does entail the conclusion that a subject has most reason to cease token evaluation of actions and to treat the authority’s commands as reason-giving, the more direct conclusion of an argument for legitimate authority is simply that the authority’s commands are reason-giving.

If one accepts the NJT as a justification for legitimate authority, then one way to answer the DA Problem on behalf of an INT is to claim that, like in the case of political authority, the structure of an indirect view grants normative power. One might say, provocatively then, that dispositions or deliberative strategies can have authority over an agent, such that the ‘commands’ of a disposition – say, truth-telling of an honest person, expressing courage of a brave person – are in fact reason-giving. This would be an interesting option, and perhaps some will find it attractive.

I do not, however, want to pursue such an argument here. In line with the ‘modest’ overall strategy of this dissertation, I do not want to solve a problem by assuming both that there are genuine normative powers, and that Raz’s argument successfully justifies such a power. Instead, I want to extract a more modest lesson from the NJT, by asking how it might help to solve the DA Problem, even if we don’t assume that it justifies a normative power.

Let us revisit the NJT, then. Raz claims that it is the normal way that a legitimate authority is justified, which means that the conditions laid out are what justify the fact that some agent or institution has the normative power of creating reasons with commands. Since my argument is not about such power, though, and I do not want to assume it for my purposes, let us
look just at the indirect structure relevant to the authority discussion. In the case of a normally justified authority, an agent would do better at responding to her own reasons were she to cease token evaluation of those reasons, and to simply attempt to obey the authority’s commands. What I want to point out is that, even if we don’t take such a scenario to imply that an authority gains a power, the following still seems to be true: the agent ought, therefore, to cease token evaluation of her reasons and to instead attempt to obey the authority’s commands. Why? For straightforward, indirect reasons: because she will better respond to her independent reasons if she does. And so in a sense, it will be the case that the authority’s commands are ‘reason-giving’: not because the authority has the power to create new reasons, but because the agent has an antecedent reason to obey the authority, which requires that the agent take the authority’s commands to be reason-giving.17 The purported authority’s power, so understood, may be uninteresting in the context of legitimate authority (since the normative force of the commands is derivative), but it is interesting in the present context.

If the above is true, then we have discovered an INT that solves the DA Problem: an agent has most reason, not merely to adopt some disposition (perhaps what I called that of being a ‘subject’ earlier); she has most reason to cease token evaluation of her reasons, and to obey the

---

17 One may wonder why having a reason to take some command to be reason-giving entails that there is a reason to obey the command. But this is not, in fact, mysterious. It is not the taking of a command to be reason-giving that makes it reason-giving; it is the reason one has to take it to be reason-giving. Consider the analogy from parenting: if I tell my child, before leaving her with a babysitter, that she is to treat any rules given by the babysitter as if they were rules that I gave her, there is a sense in which we still have to describe the situation long-hand: my daughter’s job is to take the babysitter’s rules to be authoritative, because of my authority. However, because I am giving my daughter a reason to treat the babysitter’s rules as reason-giving, they do give her reasons. The babysitter’s rules, we might say, inherit the normativity of my authority.

197
authority’s commands. And she cannot comply with this reason in any other way than by in fact obeying the authority’s commands. The strategy of obeying an authority, then, has an indirect structure that does not face the DA Problem. My next suggestion, then, is that we can model a Hobbesian argument for HCA that works in the same way.

The Razian solution works to avoid the DA Problem because it does not replace first-order reasons to act with only first-order reasons to adopt a disposition to act (which is what HCA and Indirect Consequentialism did); instead, it employs second-order reasons not to consider some reasons, and instead to obey the command (that is, to act for the reason that so acting was commanded). Thus, while we might summarize the point by saying (as I did earlier) that, when the NJT holds, one ‘ought to become a subject’, this language of disposition can be specified to show what such a strategy includes. Yes, part of becoming a subject is the development of standing attitudes and dispositions, but that is not how normativity is transferred to the evaluation of acts. One has a reason to act as a subject because the benefit of obeying orders is found only in actually obeying, and so the reason to become a subject includes a reason

---

18 As I am using the term, then, second-order reasons are just any reasons that take, as their object (that for which there is a reason) the response to some other reason. Raz, notably, focuses on one kind of second-order reasons, which he calls ‘exclusionary reasons’, or reasons “to refrain from acting for some reason” (1990b, 39-40). So formulated, exclusionary reasons are a good example of the kind of reason that I will find helpful. However, Raz is not always clear on whether exclusionary reasons can be so easily understood. For instance, in some places it sounds as though he takes an exclusionary reason to be doing something existential – that is, to be eliminating the reasons to do a thing (1990b, 40; 185); but as my use makes clear, it also sounds psychological: it is about what an agent responds to in deliberation. In the same way I avoided taking on Raz’ claim about normative powers, I will also avoid any strong read of second-order or exclusionary reasons. In my use, a second-order reason is a strategic reason, about how one should reason. As will become clear in the last half of this section, I do not think that other first-order reasons disappear, just because an agent shouldn’t consider them.
to deliberate (and not) in certain ways, and so to act on particular reasons. And whereas ‘being a subject’ could be successfully discharged even without perfectly acting as a subject, one cannot correctly respond to a reason to cease deliberation and take a command as authoritative without ceasing to deliberate and obeying the command. The inclusion of second-order reasons into the strategy of an INT reveals what, on reflection, should have been obvious: that the good of a strategy can sometimes be obtained only by following the strategy, and so the reason to adopt such a strategy includes the reason to deliberate (and as a result, act) according to the strategy.19

The move from the context of political authority to the Hobbesian argument is fairly straightforward. HCA has argued that one does better at satisfying her desires by abandoning any attempt to constantly satisfy her desires, and instead becoming a justifier, or one who is disposed towards – has a settled commitment to – living a justifiable life. The Razian lesson is that we should not leave the claim at this point, as if just having a psychological disposition towards acting in some way will confer a justification upon that action. Instead, we should take the strategy of adopting a disposition to include whatever deliberative features are justified by the move to that strategy. And, just as in the case of authority, the value of becoming a justifier is partially to be found in what one gains by ceasing evaluation of her options, and instead acting as a kind of person. The second half of the argument against the Fool was exactly that attempting to determine for oneself when to act as a justifier, and when to act otherwise, is less

---

19 It should be noted that the ‘sometimes’ here is important, as it is possible for there to be a benefit obtained simply by adopting a strategy that one never acts on. Importantly, both Schelling’s Answer to Armed Robbery and the problem of the threat-enforcer are like this: in such cases, the benefit of a disposition is only in having the disposition; in fact, acting on the disposition is disastrous. This distinction will be important in my final evaluation (in 5.8) of whether all indirect strategies generate reasons to follow the strategies.
likely to be successful (for creatures like us) than is simply acting as a justifier.\textsuperscript{20} This means that we can treat dispositions as having ‘recommendations’ (the disposition towards justifiability recommends acting justifiably, the disposition of covenant-keeping recommends keeping one’s covenant, etc.), and the reason to adopt such a disposition often comes with the reason to take those recommendations as ‘authoritative’.\textsuperscript{21} Metaphorically, we can help to see the structure at work by saying that we have most reason to be justifiers, because ‘obeying the dictates’ of such a disposition is likely to best satisfy our desires. And so, since the value of being a justifier is grounded in deliberating and acting in a certain way, one has most reason to deliberate and act in that way – namely, one has most reason to cease token evaluation of what will satisfy one’s desire whenever being a justifier recommends some action, and to simply obey that recommendation.

The Razian argument above partially solves the DA Problem for HCA. The solution is only partial because it explains how one gets from reasons to adopt a disposition to reasons to act, but it does not eradicate any countervailing reasons to act that already exist. Thus, in those (most interesting) cases where an indirect strategy recommends an action opposite what would be recommended by the token reasons, we are left with conflicting reasons that are grounded in very different kinds of considerations.

\textsuperscript{20} Recall that the first half of the argument was simply that particular desires are best promoted by \textit{having} a certain disposition. These were the two methods discussed in 5.2, particularly f. 5. \textsuperscript{21} Notice that, at this point, my argument may seem to have made sense of the Gauthier claim that I earlier criticized: it \textit{does seem right} that, if one ought to adopt some disposition, then one ought to act on it. However, his argument is for the strong ‘middle premise’ that allows one to derive B from A, whereas my argument is only for a moderate claim that we can \textit{sometimes} derive B from A. Thus, I have not completely vindicated Gauthier, although I have provided an argument that explains why his claim often seems true.
Consider the following example. Ercan has worked long and hard on an unpaid internship alongside 30 other young people. At the end of the internship, only one of the interns will be hired into an amazing job, and the decision will be made on the basis of a competition, in which each intern presents a proposal to the CEO, who will then choose the best one. As the competition nears, Ercan realizes that his proposal is simply not very good, and that he doesn’t stand a chance at winning. However, a friend, Sarah, who is on the Board of Directors, offers to save him by passing along a proposal that she knows will be attractive to the CEO, which he can then submit as his own. Sarah’s proposal hasn’t been seen by anyone, and she can virtually guarantee that presenting it will land Ercan the job. And since Sarah is the single most loyal friend Ercan has ever had, he has no worries that their secret would ever come out.

If Ercan is sufficiently like me, my friends, family and colleagues, then he has a strong desire to live a justifiable life, and a very strong desire for intimate relationships. Assuming the arguments of this project, then, Ercan has weighty reason (and likely most reason) to be a justifier. As a justifier, he likely is habituated to turning down offers such as Sarah’s, and may even do it automatically from his standing commitment to the standards of justifiability; however, those psychological features do not explain the relevant reasons. Instead, what we can now say is that Ercan has most reason to treat the recommendations of ‘being a justifier’ as authoritative – that is, he has most reason not to evaluate, or not to act on his evaluation of, the independent considerations of the case, but instead to simply obey his disposition’s recommendation to act justifiably.22 And presuming, what seems plausible in this case, that it

---

22 Notice that Ercan has *most reason* to act on his disposition’s recommendation, rather than merely ‘some reason’, because he has most reason to be a justifier. The argument for solving the
would be unjustifiable to pass off Sarah’s proposal as his own, then Ercan has most reason to decline her offer.

This does not mean, however, that Ercan has no reason to accept Sarah’s offer. Clearly, he has many reasons, and likely weighty ones. That he had most reason not to consider, or not to act on, those reasons does not imply that the reasons in fact disappeared. And so our ‘moral reasons’ – those in favor of acting justifiably – do not ‘silence’ or ‘undercut’ competing reasons on this view. Instead, we are left admitting that the normative reality is messy, as Ercan’s case demonstrates.

My final lesson, which I will draw from Railton’s contribution to the topic, is that this is how I think we should want the view to turn out. If Ercan turns down Sarah’s help, then we will say that he did as he had most reason to do, and perhaps (with the aid of some additional arguments and connections that I will not make here), that he did what he ought. But if he accepted Sarah’s help, it is still true that he acted in a way favored by weighty reasons. Indeed, if we knew more about his struggles and goals, then without changing our judgment on whether

---

DA problem was not that there was some normative transfer between dispositions and actions; but rather, that one has most reason to be a justifier, and that part of the strategy of being a justifier is obeying its recommendations. That is: at least in the case of this particular disposition, the desires that ground reasons to adopt the disposition also ground reasons to ‘take the disposition’s recommendation as authoritative’.

23 To be clear, then: since Ercan has most reason to cease evaluation of his token reasons, and to act on his disposition of being a justifier, he has most reason to turn down the offer. However, that he has most reason to deliberate in some way and to act in some way does not imply that weighty, but outweighed, reasons cease to exist. Thus, Ercan may have most reason not to respond to reasons that would have, in the absence of his reasons to adopt the strategy of being a justifier, been the weightiest reasons to act. Which is just to say: if one has most reason to be a justifier, then he has most reason to ignore the considerations that make it the case that he has most ‘token reason’ to act counter to his disposition.
it was justifiable, we might say that his mistake was an understandable one, as the reasons favoring the ruse were so compelling. Because second-order reasons – as I have been using them – do not silence the competition, there will often be normative residue in the aftermath of an act, as there is the potential for genuine normative force on the losing side of what one has most reason to do.

This insight is borrowed from Railton’s discussion of Indirect Consequentialism, where the point is even easier to see, I think. Railton claims that, if his argument for becoming a sophisticated consequentialist works, then it will likely entail that one ought to develop stable character traits, habits and dispositions that will lead one to act ‘wrongly’, given consequentialism’s criterion of wrongness. Because Railton did not attempt to generate ‘rightness’ or reasons to act from the rightness and reasons to develop a disposition, he is left with an even muddier situation than I am, as he is left saying simply that I perhaps ought, morally, to become a good husband, and that this may reliably lead to my acting wrongly (when I spend too much on my spouse instead of sending it to Oxfam, for example) (1986, 120-121).

As I will discuss briefly in closing, I think Railton could also utilize my Razian argument to say something a little more satisfying here. However, the intuition that I think he makes good sense of concerns the genuine normative conflict in those situations where the grounding reasons and the strategy for following them come apart. These cases shouldn’t come to a clean conclusion, because the cases aren’t clean: I should be a good husband, and as I think (although Railton didn’t argue), I should even spend more resources than necessary on making my spouse feel loved and special. But that doesn’t mean that there aren’t very good, very compelling reasons to do something else with those resources, and the existence of those reasons calls for an
experience of loss when I cannot respond both to them and to the reasons to be a good husband. Indirect Normative Theories result in the possibility (and actuality) of normatively messy cases, where the reasons to follow a strategy are opposed by reasons for which one adopted the strategy in the first place. But this is not a cost of the theory, as it seems to accurately reflect reality in the relevant cases.

5.8 Schelling and Threat Enforcers, *Redux*

At this point, I have argued that, for INTs like Raz’ account of authority, Railton’s consequentialism, and my own HCA, the first-order reasons for adopting a disposition come, as part of the strategy, with second-order reasons to cease evaluation of one’s token reasons, and to act on the recommendations of the disposition. A natural worry, then, is that my argument has proved too much, as it looks like I have shown precisely what Parfit and Skyrms argued couldn’t be shown, which is that those who have most reason to adopt certain, destructive dispositions thereby have reason to act on those dispositions. But wouldn’t it be a strike against the argument if it proved that Schelling in fact had most reason to beat his children, and a country had most reason to launch a catastrophic nuclear attack?

Recall, however, that I have not argued in favor of some middle premise that automatically generates B from A (the claim that one has reason to act as a justifier from the claim that one has most reason to be a justifier). If I had, then the conclusion of my argument would be something like the following: one ought to act on the dispositions that one ought to have; or in Gauthier’s language, if a disposition is rational, then surely so is acting on that disposition. And in this case, I would, in fact, be committed to the implication that Schelling ought to beat his children, and that the nuclear threat-enforcer ought to launch a catastrophic nuclear attack?
attack. But I am not so committed. In this final section, I will argue that the claims of this chapter so far provide a natural explanation of why the nuclear threat-enforcer has most reason not to launch, and why Schelling likely has most reason not to beat his children.

Note that the key move, borrowed from Raz, in solving the DA Problem is to highlight that in the cases here discussed, *it is the following through on the strategy* that makes the strategy good. That is: one best responds to her reasons *not merely* by having the disposition of a subject, or justifier, but by adopting the strategy of ceasing to evaluate her token reasons, and instead taking the dictates of the authority (or the recommendations of the disposition) as authoritative. In the cases of authority, HCA, and indirect consequentialism, that is because part of the justification for adopting the indirect strategy is due to inconsistency and unreliability of the agent’s own evaluation of the reasons. The Razian move is to note that, *in some cases*, the reasons one has to adopt a disposition come with reasons to ‘take the disposition as authoritative’.

Now we must simply note that not all strategic reasons to adopt a disposition are like this. Some are not based on the inconsistency and unreliability of the agent, but based *only* on the value of *having* the disposition. The case of threat-enforcement, particularly in the one-off, nuclear catastrophe version, seems like a good candidate for such a disposition. If my best strategy for ensuring global survival is to become a threat-enforcer and then threaten a rogue nation with nuclear war if they do X, then I have most reason to do just that. However: the good of being a threat-enforcer here is not like that in the case of HCA or Raz’ political authority. One ought to adopt the disposition of being a threat enforcer because doing so is most likely to promote global survival; however, it is not the case that the value of being a threat enforcer is in
ceasing evaluation of one’s token reasons – the value is simply in being seen as a threat-enforcer. Thus, the Razian argument does not justify the inclusion of second-order reasons into the nuclear threat-enforcer’s strategy. What is best, in that case, is to become a threat-enforcer, but specifically not to cease evaluation of one’s token reasons. Now, as a matter of fact, it may turn out that becoming a threat-enforcer has that implication, but this psychological fact will not confer justification on launching when the time comes. If one had only first-order reasons to adopt some disposition, then this does not imply that she thereby has reason to act on that disposition. And so the nuclear threat-enforcer, although she may, as a result of psychological force, hit the launch button, has no disposition-based normative reason to do so, and overwhelming reason not to.

The case of Schelling, then, may of course be similar, depending on the scenario. If the only value in becoming irrational is being seen as irrational by the burglar, then Schelling will be like the nuclear threat-enforcer: he will have most reason to drink the serum, but most reason not to act on it. However, if one spells out the Schelling case in a way that makes it more like the INTs discussed above, then our judgment may (and may reasonably) change. If part of the justification, say, for drinking the serum is that one is likely to misjudge how severely he must act, then the recommended strategy will indeed incorporate second-order reasons: Schelling, in such a case, ought to drink the serum because he will do better at responding to his reasons by ceasing evaluation of token reasons and instead to simply obey the dictates of ‘irrationality’.

I have not, in this chapter then, argued for a middle premise that necessitates moving from A to B. It is not the case that one always ought to act on the disposition that one ought to have. Instead, I have argued that some indirect strategies involve adopting dispositions that one
ought to ‘take to be authoritative’; that is, that sometimes one does best not merely by adopting some strategy, but by acting on it. As a result, a Razian account of ‘being a subject’, HCA’s account of ‘being a justifier’, and Railton’s account of being a ‘sophisticated consequentialist’ are each able to solve the DA problem, but without the implication that the nuclear threat-enforcer ought to initiate global annihilation.

5.9 Conclusion: Putting the Solution to Work for Other Indirect Theories

The arguments of this chapter are many, and the conclusions are some of the most ambitious of this dissertation. I have shown that the Problem of the Fool requires that HCA move to a claim concerning what dispositions one has most reason to adopt, and that this constitutes making HCA an Indirect Normative Theory (INT). And while the move to an INT is seductive in response to the Fool, making it invites the Disposition-Act Problem (DA Problem). Fortunately, HCA need not solve the problem by itself, as two prominent INTs each make partial headway towards solving DA Problems.

The positive proposal, then, is this: Raz’s argument can be used to branch the disposition-act gap, and Railton’s argument can be used to assess the situation we are left with. I think that this is a promising solution, not only because it solves the Fool-DA Problem for HCA, but because it provides a model of solving versions of the Fool-DA Problem that occur for every INT. In drawing from Raz’ argument, for instance, I modified it in a way that makes sense of how an authority’s commands provide reasons (but does not create them) without recourse to contentious assumptions about normative powers, utilizing only the indirect structure of the theory. And in drawing from Railton’s insights, I noted that he was required to stop short of explaining reasons to act on the dispositions that it is right to have, but my Razian argument
could bridge that gap. In short, the discussion of INTs here could be used to draw the strengths from each individual theory, and apply it to any other. While my current interest is in applying the lessons to my own view, HCA, I believe that an Indirect Consequentialist could do the same.

But I leave completing that task to the consequentialist. If my arguments here have met with any success, then I have removed what may be the biggest obstacle to accepting a Hobbesian view of any kind, my own included – and that is the Problem of the Fool. I will now turn to my final defensive chapter, in which I respond to a collection of worries that HCA is ultimately unsatisfying, as it fails to vindicate too many of our considered moral judgments.
CHAPTER 6

A Humean Contractualist Account of Moral Reasons (HCA) is, I have argued, promising in a variety of ways, and defensible against the very difficult problem of the Fool. If so, then the arguments of this dissertation serve to sketch a new Hobbesian theory that deserves at least some initial consideration, as it claims to generate a powerful explanation of moral reasons from philosophically humble beginnings. Unfortunately, many philosophers are likely to hold that initial consideration is all that is required in order to reject the view, as it looks to have serious problems. Although the challenges that I will address in this chapter are several, they share a common core.

There is a tension running through this dissertation, which is common to all minimalist explanations of morality. In short, it is that I appeal to the kind of explanation that is not typically thought well suited to explaining moral reasons (desires of particular agents), and then argue that we actually can generate a plausible explanation from such humble grounds. And any view that takes such a tack must be aware of danger on both sides of the project: on the one hand, claim to account too well for the moral phenomena, and no one will believe the conclusion. Whatever ‘magic bullet’ argument is invoked to turn bare self-interest into moral restrictions will be the subject of incredulity. Give up on perfectly explaining ‘our’ moral intuitions, though, and that will also be seen as reason enough to reject the view. Certain conclusions seem to many to be non-negotiable theoretical desiderata of a moral account: morality must be overriding and reason-giving, one might challenge, or the moral truths must be non-relative.
Over the course of the previous 5 chapters, I have attempted to give a relatively satisfying explanation of moral reasons, but from philosophically modest grounds. In doing so, I have invited criticism from both sides of this tension. Since I am suspicious of all extant ‘magic bullet’ arguments, which claim to pull universality out of the minimalist hat, I have accepted some amount of modesty in my conclusions: very many of us have most reason to act rightly. And this is sufficiently weak to raise worries that HCA is extensionally incorrect. On the other hand, I do not think that every person’s reason-set is completely idiosyncratic, which is why I can conclude even as strongly as I do – that very many of us have most reason to act rightly. And this conclusion required my own ‘weak’ magic bullet argument, in which I claimed to show that a wide-spread desire for intimacy can generate a weighty reason, which is surprising enough to raise the incredulity worries.

The various challenges raised here bring to light what my opponents have likely been thinking all along – namely, that HCA is not a real candidate theory of moral reasons. It simply has too many costs. And it is for precisely this reason that I have invited the reader to remain open-minded by considering it as a ‘back-up’ view of moral reasons. After all, if one’s preferred account ultimately fails, would it not be better to have some confidence that many of us, nevertheless, have most reason to be moral? But if the challenges previewed above are moving, then even if I have made a case for HCA being a promising backup view, my opponents likely feel vindicated in not according it the standing of being a *primary* candidate view. HCA, it is being objected, simply has too many costs to be taken seriously as a contender for being one’s preferred view.
In this final chapter, I will do my best to argue that the facts of the matter concerning the costs of HCA are not nearly so clear. I will do this by addressing the concerns on both sides of the tension raised above, although I admit up front that this will occasionally require conceding that HCA holds exactly what its opponent fear. Especially regarding the non-universalism of HCA, much of my response will simply borrow from others’ arguments to the effect that this is not as inadequate a conclusion as it seems. However, in defending my inference from a widespread desire for intimacy to a common, weighty reason to be moral, I have some work yet to do, as there appear to be counterexamples to the claim that there is a tight (even if contingent) connection between the ability to have intimate relationships and a commitment to justifiability. Although I claim to have an explanation of such counterexamples, the explanation invites several additional worries, which occupy me for the remainder of the chapter.

6.1 Defending Modesty: Non-universalism

Many philosophers endorse ‘morality/reasons’ internalism (or, for ease: moral internalism) – the thesis that the dictates of morality necessarily provide agents with normative reason to act. I take the label of moral internalism from Darwall, who distinguishes the view from judgment internalism (the view that sincere moral judgments necessarily motivate), and reasons internalism (the view that normative reasons must be capable of motivating the agent for whom it is a reason (1997, 306).1 If this thesis were true, then everyone would have a reason to do what morality commands, and so morality would be universally normative.

---

1 The view that morality is necessarily reason-giving, however, has been made under different headings by many philosophers. Richard Joyce, for instance, argues that morality necessarily has ‘practical clout’, which is to say that moral claims necessarily provide agents with categorical reasons (2006, 62), and Michael Smith argues for the view that he calls ‘moral
Such a view is attractive, and it is easy to see why: if it were false, then morality and reasons might come apart, and so there may be individuals who have no reason to do what morality commands. And this possibility is disturbing, for both forward- and backward-looking reasons. Looking forward, it seems that by convincing someone that acting in a certain way is impermissible, we are showing her that she has a reason not to do it – that is, that we are getting a normative grip on her. And looking backward: if our interlocutor ignores us and commits the forbidden action, it seems that we correctly say that she ought not to have so acted, which would seem to imply that she had a reason not to so act. If moral internalism were false, then, we would be robbed of two different kinds of important moral tools: we would both fail to get a normative grip on some individuals, and fail to truthfully condemn them in a particular way. Both implications seem like costs.

HCA, it should be clear, does not vindicate moral internalism. According to HCA, the link between the reasons that some agent has and the dictates of morality is a contingent one, as rationalism’, which holds that moral facts are facts about our reasons for action (and the rejection of rationalism, according to Smith, is ‘externalism’) (1994, 62). See also Alexander Miller for a nice discussion of all of these ‘internalist’ views (2003, 225-228).

The rejection of moral internalism may seem too strange to take seriously; surely, one might think, it is simply a conceptual truth that morality provides agents with reasons. And, as one can see from the previous footnote, those who believe this are in good company. However, it is worth noting that moral externalism is not new or undefended. The locus classicus for moral externalism is Philippa Foot’s famous (1972), but a variety of views have defended similar externalist commitments. The so-called ‘Cornell Realist’ positions of David Brink (1986; 1989) and Nicholas Sturgeon (1985) are examples of moral externalism, as is Peter Railton’s influential view from (1986). A common theme shared by each of these views, along with HCA, is that they provide one explanation for morality, and another for practical reasons, rather than beginning with the assumption that moral facts are facts about reasons, or are necessarily reason-giving. And while each view would likely be odd if it posited no connection between morality and reasons, the externalist position simply admits that whatever connection exists will be
whether or not one has a reason to be moral depends on the content of her desires. Even if it were the case that everyone alive today happened to desire to live justifiably or to be in intimate relationships, the morality/reasons connection would be merely contingently universal. In that case, my opponents would worry that, even if everyone alive had reason to be moral, the nature of the connection allows for the possibility that there could exist someone who had no reason to be moral. And that, for some, would be sufficiently worrisome.

HCA, then, both fails to forge a necessary connection between reasons and morality, and it fails to claim a contingently universal connection between reasons and morality. Thus, according to HCA, it is clearly possible, and likely actual, that there are some individuals who have no reason to act justifiably. This is at least a disturbing implication, and some will say a reductio ad absurdum of the view. In what follows, I will argue that the failure to vindicate moral internalism, while perhaps a cost, is not devastating to HCA.

6.2 What We Can and Cannot Say to and about Hitler

I noted above that moral internalism preserves our ability both to normatively grip others, and to accurately condemn them in a particular way. A concern with HCA, then, is that it will fail to do both of these things in important cases. In the case of moral monsters, for instance, it will be objected that surely we are correct to say that they have reason to do other than they do, both prospectively and retroactively. Surely, that is, we are correct in our judgments of Hitler—that he did monstrous wrong, and that he shouldn’t have done what he did. Now, since HCA adopts Scanlon’s contractualism, it can say that it is simply true that Hitler acted wrongly, as it is contingent, and so there may well be particular instances in which moral facts do not entail practical reasons for some agent.
certainly the case that many of his actions were unjustifiable. But the real worry here is that such a claim is impotent if it doesn’t imply a reason to act in accordance with rightness. In fact, if Hitler’s desires were lined up in a sufficiently sadistic and monstrous way, then he had most reason (according to the proportionalist Humeanism of HCA) to do as he did. And that seems like an unacceptable conclusion.

There are multiple cases to be investigated here. The invocation of Hitler has a particular urgency to it, as he was a real person, and so his evil is particularly terrifying. But dealing with that actual case would mean not making too many unsubstantiated assumptions about Hitler’s psychology (like the assumption that he had uniformly evil desires). And so I will discuss Hitler first as a (fairly) realistic case, and note what it is, in fact, that we can and cannot say about him and his actions. I will then move on to the second concern, that there could exist a pure case of moral monstrosity, and so will reflect on Gibbard’s example of the ideally coherent Caligula.

In the first case, the actual Hitler was almost certainly – as we all are – a complex person. There seems to be good reason to believe that he had terrible desires – to purify the Aryan race, for instance – as well as a willingness to pursue that desire over many other, what one would expect would be competing, desires. After all, the Third Reich cost Hitler much, ultimately including his life. And, if he was anything like other people, he probably desired relationships, success in his projects, and to survive. Although hard to believe, he may have even had an

---

3 However, the separation of wrongness from reasons may render more urgent the question of exactly what is added to the observations about Hitler when one says that he acted wrongly. That is: since wrong action just is unjustifiable action, and unjustifiable action isn’t reason-giving in this case, why should we care whether or not the action is appropriately described as wrong? How does such a label help? I address these questions in the following, concluding chapter.
intense to desire to live justifiably, as it is possible that he believed his actions were, in fact, commanded by duty. Of course, if he believed this, he was mistaken, as the contractualist layer of HCA establishes; but the question of his reasons, according to HCA, depends on what he desired, regardless of whether he was correct about how to satisfy those desires.\(^4\)

Note, then, that ‘the’ case of Hitler actually presents us with multiple, possible cases, requiring multiple, different responses. If it were true, for instance, that Hitler had an intense desire to live justifiably, then the case presents a different challenge from the one we are currently investigating. On that scenario, he clearly had weighty reason to do other than what he did, because he was in serious error about how to act justifiably. The burden for HCA in that case is to provide some explanation of the standard of justifiability that shows it to be sufficiently non-relativist; I will touch on these worries in the latter sections of the chapter.

The other cases of Hitler, then, provide us with a much messier picture. If the arguments of this dissertation are on target, then Hitler had a reason (defeasible, of course) to act justifiably so long as he had a desire for intimacy; and he had a reason to do other than he did generally if he desired that things go differently than he had reason to think his actions would lead them to go. It is plausible that both of these conditions obtained. Hitler, like many human beings, may well have desired deep friendship and romantic love, and if he did, then he had a reason to act justifiably. He also likely had a desire not to die young – his armies decimated by the Allies –

\(^4\) Recall that the Humean view imported in this project holds that R is a reason for A to phi if A desires that X, and the truth of R at least partially explains how phi-ing promotes X. And notice, now, that there is plenty of room for all of us to be in error concerning our reasons. If, as seems obvious, abandoning the Third Reich best promoted living justifiably, and if Hitler did desire to live justifiably, then Hitler had a reason to abandon the Third Reich – even if he \textit{believed} that purifying the Aryan race was commanded by duty.
and so depending on the forward-looking plausibility of his success, this desire also might have provided him with reason not to instigate a World War.

The case of Hitler was messy, then, but what still seems plausible is that his warped, evil desires (however they are spelled out – maybe for the eradication of Jews and other minorities) were strong. And this is precisely why many find proportionalism to be the downfall of Humeanism: it would thus seem plausible, even on the messy, realistic case of Hitler, that he had weighty reason to do as he did; perhaps he even had more reason to commit genocide than to do otherwise. And so we are robbed of our ability to say what seems obviously true, that Hitler had most reason not to lead the Third Reich and to commit genocide.

Even in this case of Hitler, there is much that I can say, according to HCA. Plausibly, I can say that he was a monster, that what he did was wrong, that it was good that he was defeated, and likely even that anyone who acts as he did ought to be blamed and to face severe penalties. There is a sense in which I can even say that Hitler clearly ought never to have done what he did. In fact, there are very few things that I cannot, truthfully, say in condemnation of Hitler, other than ‘that he had most reason to act differently’ and ‘that he had no reason to do what he did’. In the remainder of this section, I will explain how it is that we can and cannot (truthfully and sincerely) say these particular things, and argue that such an explanation shows that we are not really limited much at all in reference to how we may criticize Hitler.

It is clear that what Hitler did was wrong, according to HCA, because HCA holds that Scanlon’s contractualism is true, and Hitler’s actions certainly were not justifiable. In addition, it seems plausible that character judgments might come apart from ascriptions of reasons, such that someone’s being a ‘monster’, or even a ‘bad person’ need not imply that such a person fails
to respond to the reasons that apply to him. Indeed, a thick concept like ‘monstrous’ likely has many nuances and contains both descriptive and evaluative components: perhaps (merely for illustration) one is monstrous by showing complete disregard to the significant harm that he is causing others, given that we – the rest of us – take the fact that an action will harm others as a reason not to do it. Such a view of monstrousness combines purely descriptive facts with facts about what some people – at least observers – take to be reasons, and so makes good sense of the ‘thickness’ of the term. While I will not here defend a particular analysis of ‘monstrous’, I will just state that any account like the above, which combines descriptive elements with facts about what at least some of us take to be reasons, seems to apply to someone like Hitler.

For similar reasons concerning a distinction between the evaluative and deontic, there are plenty of evaluative claims of the form, ‘it is good that Hitler was defeated’, that don’t imply that Hitler had any reason to do other than he did. It was clearly good for the Jews, for other minorities, for the millions of soldiers, and, in short, for most people of the world. It is thus at least clear that we could be using ‘good’ in a relative sense (‘good for’); however, it is also possible, depending on one’s value theory, that it is simply good that Hitler was defeated. Again, I will not here provide a value theory to justify this claim, but it seems many views would be able to accommodate it without implying anything about Hitler’s reasons.

The case of what Hitler ‘ought to have done’ is more interesting, and here I borrow from Gilbert Harman (1975). In defending a form of moral relativism, Harman takes care to explain that our ‘ought’ claims are multiply ambiguous, and that many of them can truthfully be made even if moral universalism is false. Although his argument is thus slightly different from my own (his view may well be moral internalist, but the moral claims themselves are not universal),
the structure is similar, and so many of Harman’s observations transfer over to the evaluation of HCA.\textsuperscript{5} For instance, Harman holds that one common use of ‘ought’ is in the context of evaluating a situation: so Hitler ought not to have committed genocide; while one reading of this claim ascribes reasons to Hitler, another doesn’t. One thing we might be saying, according to Harman, is that it ought not to have been the case that Hitler did what he did (Harman 1975), and perhaps even that others ought to have stopped him.\textsuperscript{6} The Third Reich was terrible, and it is likely that many people and institutions ought to have done something about it.

In fact, Harman goes on to say that the one claim that he can’t make about Hitler – what he calls the ‘inner ought judgment’, or the normative judgment that would carry some deliberative weight with him – would sound ‘odd’ if we were to say it. It is the wrong kind of claim to make about Hitler, because he was a monster. He was ‘beyond the moral pale’, so to speak, and ought-claims seem to be the kind of thing that provide an agent with a reason to which they might respond. Thus, according to Harman, the one claim that his non-universalist view cannot vindicate is one that in fact, shouldn’t be vindicated (1975, esp. pp 5-8).

I don’t know how convincing these suggestions by Harman are, but however promising they are for his own view, they are even more so for HCA. Since HCA does not take a stand on how ‘ought’ is to be analyzed, it is not even clear that we can’t say that Hitler ought not to have

\textsuperscript{5} That is: although Harman was a moral relativist, many of his arguments are helpful to my own, non-relativist position. The goal that he and I share is of explaining how we can truthfully make certain, intuitive claims, despite the vindication of a counter-intuitive view. The counter-intuitive view, in his case, is a first-order moral relativism, which holds that the criterion of right action depends on cultural context; the counter-intuitive view for HCA is that some moral monsters have reason to act as they do, despite the fact that it is wrong for them to so act.

\textsuperscript{6} This further implication is attributed by Harman to Thomas Nagel (Harman 1975, 6 f. 2).
done otherwise.\textsuperscript{7} What we \textit{do} know, however, is that we cannot say that Hitler had most reason to do other than he did, or that he had no reason to act as he in fact did. Anyone who shares Harman’s intuition about the oddness of ought ascription in the case of Hitler, then, should be even more confident in the case of reasons-ascription. \textit{Because} Hitler was such a monster, it is implausible (says Harman) that our reasons could ‘get a grip’ on him. Or, from the other side: \textit{of course} it’s false that he had no reason to act as he did: \textit{he was evil}, and evil brings with it its own kind of reasons.

To be clear, I do not think Harman’s argument is very successful against either his or my own strongest opponents. It relies on the very intuition that many will deny – that normativity must be able to grip an agent \textit{from the inside}. On such a view, whether I have a reason or ought to do something depends on my having a certain psychology, or having commitments that might allow me to understand those considerations as reason-giving. What the argument \textit{does}, however, is show that the class of claims that cannot be truthfully said on a non-universalist view is both narrow, and plausibly (for some) exactly that class of claims that are less obviously true anyway.

Before closing this section on what we can and cannot say about Hitler on a view like HCA, I want to consider for a moment the \textit{perspective} of many of the claims we make. So in addition to the arguments and suggestions above, one might point out that when we make

\begin{footnotesize}
\begin{itemize}
\item That is: one view holds that ‘ought’ simply means ‘has most reason to’. If this is the correct view, then clearly HCA cannot say that Hitler ought not to have acted as he did (in the ‘internal’ sense). However, one might think that what one ought to do, at least sometimes, is dependent on one’s obligations, \textit{independent of one’s reasons}, in which case it would be true even that Hitler ought not to have committed genocide.
\end{itemize}
\end{footnotesize}
normative claims, we might also just be using a kind of implicit domain restriction or indexical; in other words, our claims about Hitler might not be about Hitler himself, but rather, about what is true of a person like Hitler according to one’s own standards. As an example: moral relativists and Humeans, like most of their friends, family and colleagues, may believe that they can make clear moral and normative judgments of people in the real world. Upon hearing about a brutal rape, such a person, I would imagine, issues the most ardent condemnations of the rapist, and believes all such claims to be true, regardless of the psychology of the rapist. On the relativist’s picture, this may be because the speaker is condemning the actor from within the speaker’s set of norms; and on the Humean picture, it may be because, whether or not the rapist is in fact anything like a psychologically-normal person, we each judge him from that perspective. ‘Our’ set of reasons (those of us who care about some shared set of goods, like family, intimacy, stability) clearly outlaws this action.

While I think that such an explanation actually makes good sense of many of our actual claims, some find it disturbing because of the connection between such claims and the practices of blame and punishment. The worry is that blame and punishment must be justified, and that there is something horribly unjustified about blaming and punishing someone for doing what they had weighty reason, and maybe even most reason, to do. This brings us to the question of whether we can vindicate claims concerning the appropriateness of blame and punishment on a picture like HCA.

Although I certainly cannot provide complete accounts of blame and punishment here, I want to point out that attractive views are compatible with HCA as laid out above. Scanlon’s own theory of blame, for instance, is able to make good sense of how Hitler’s actions could be
blameworthy, and why we would in fact blame him, even if he had most reason to act as he did.\(^8\) This is because, according to Scanlon, we are not concerned only with whether an agent acts according to reason,\(^9\) but also with the agent’s intentions – more particularly, with the reasons to which the agent responds.\(^10\) When we blame someone, then, we modify our relationship with that person in light of how they respond to reasons – which reasons they respond to, and which reasons they ignore – and if so modifying the relationship is appropriate (whether or not it’s actually done), then the action is blameworthy (2008, 141). There are clearly many details that must be investigated here, and Scanlon addresses most of them in developing the view.\(^11\) For my purposes, however, it is sufficient to show that a prominent view – one that has, in fact, been paired with contractualism in the literature – can make sense of our judgments of blameworthiness even in cases where the target of the judgment acted in accordance with reason.

I will have the least to say about the case of punishment-claims, as we are moving further away from central, moral-theory concerns and beginning to intersect with broad questions in political philosophy and public policy. However, it is worth noting that at least one prominent theory of punishment can straightforwardly make sense of the claim that Hitler ought to be punished, even if he acted in accordance with his reasons – namely, consequentialism. In fact, consequentialism may seem more plausible with reference to punishment than in the moral

---

\(^8\) See further the brief discussion of Scanlon’s theory of blame in the concluding chapter.  
\(^9\) Scanlon calls actions that are in accordance with reason ‘permissible’ (2008), but obviously I cannot take on this terminology. Since I have denied moral internalism for HCA, an agent’s having most reason to do something does not imply that the act is permissible. So here, I will simply use the language of ‘accordance with reason’ to refer to Scanlon’s category of permissibility.  
\(^10\) Scanlon calls this the ‘narrow sense of intention’ (2008, 10).  
\(^11\) See further ch. 4 of (Scanlon 2008).
theory context, as it allows us to avoid what many (myself included) see as the indefensible aspects of more retributivist theories of punishment. It seems plausible that Hitler ought to be punished, even if he was acting in accordance with reason, because failing to punish mass-murdering tyrants would have bad consequences. Because consequentialism is ‘forward-looking’ rather than ‘backward-looking’ – i.e. it justifies punishment by appeal to what will happen as a result, rather than by what was done – it implies no necessary connection between punishment and acting according to one’s reasons, and it avoids the problems of retributivism.

I am sure that there are many other theories of punishment that would work as well, but a single example of a prominent view will suffice for showing how HCA might work. And again, by adopting some kind of forward-looking theory of punishment, HCA remains friendly to its Scanlonian roots. While I do not know that Scanlon has anywhere embraced a particular theory of punishment, he has made explicit that he rejects the Desert Thesis, which holds that it is good that bad things happen to those who have acted wrongly (1998, 274). If, along with Scanlon, we reject the Desert Thesis, we are saying that while it might be necessary (to protect innocents, for instance) that bad things happen to people who have done wrong, it is not good that such things happen. To claim otherwise seems, to the non-retributivist, explainable only by reference to our ugly desires for revenge. It is thus not only acceptable that HCA breaks the connection between punishment and acting in accordance with reason, but it may in fact be necessary in order to conform to a defensible theory of punishment.

If what has been sketched above is plausible, then the threat of the Hitler case is not nearly so disturbing as it might originally seem. We are not robbed of our ability to say most of what we want to say; and the few things that we cannot, in fact, truthfully say are plausibly
instances of things that would be odd to say (according to Harman). At the very least, revealing how few judgments are undermined by HCA ought to make us wonder how strong an objection the Hitler case really is. If HCA undermined our ability to truthfully say most of what we thought was true, then perhaps this would be a major strike against the view; however, I have shown that HCA can explain why the vast majority of what we actually say about Hitler is true and appropriate. Although some will disagree with Harman, and see HCA’s inability to account for the truth of certain claims about Hitler as a cost, it simply doesn’t seem like a very high cost. The claims in question are of a narrow class and rather technical in nature; it is difficult for me to see anyone taking HCA’s explanatory abilities with reference to this class to be decisive one way or the other.

6.3 The Ideally Coherent Caligula

Part of my response to the ‘Hitler objection’ relied on my insistence that we take Hitler – a real person – as having a realistic, messy, likely conflicted psychology. Thus, one of the things that I think we are likely to be able to say about the actual Adolf Hitler is that he in fact had reason not to act as he did. However, such a response invites the counter that, whether that was true of the actual Hitler or not, it need not have been true. Surely it is possible, my objector challenges, that someone exists who has perfectly uniform, terrible desires. So as not to confuse this challenge with the one above, I will utilize Alan Gibbard’s case of the Ideally Coherent Caligula, who has a perfectly coherent set of sadistic desires (Gibbard 1999, 145). For my purposes here, I will assume that Caligula desires only those things that give him reason to act in ways that the rest of us find repulsive – to torture, maim, rape, and murder. It appears, then, that HCA must hold of Caligula that he has most reason to do these things, and no reason at all to
refrain from doing them. It seems, in other words, that HCA must recommend that Caligula act in what we take to be horribly immoral ways.

My opponents may well take the challenge of the Ideally Coherent Caligula to be the most difficult version of the objection to non-universalism. I, in fact – following Sharon Street and Gil Harman before her – find this to be a much easier case than the Hitler one above. The reason, in short: although I admit that HCA must recommend that Caligula act in terrible ways, I also believe that Caligula is such a strange creature that we shouldn’t find HCA’s conclusion all that counter-intuitive.

The first thing to realize about the Ideally Coherent Caligula (or any of his friends – Rawls’ blade of grass counter (1971, 432), Gibbard’s ideally coherent anorexic (1990, 171), Parfit’s man with future Tuesday indifference (1984, 124)) is that they don’t exist. Of course, in a sense, my opponent realizes that they don’t actually exist, but the objection is that they could – that there is nothing impossible about such a being existing. But whether possible or not, the most coherent human is wildly far from being ideally coherent. Actual psychologies are a complete mess, and every one of our thousands of idiosyncratic desires ground some reason or reasons, according to HCA. So for a small example, consider: everything that I desire that depends on my freedom gives me reason not to lose my freedom; my desire for ice cream, then, gives me reason not to go to jail (since in jail I couldn’t choose to satisfy my desire for ice cream), and my reason not to go to jail explains one reason not to mass murder or to torture for fun. So the Ideally Coherent Caligula, if he really is ideally coherent, and really does have no reason not to torture for fun, must not desire ice cream, sex, life, a comfortable bed, or any one of another million things.
The Ideally Coherent Caligula, then, is – as Street puts it – far stranger than the case makes him seem. He is not an existing person with a few minor tweaks; he is a very different sort of creature, who just doesn’t look (psychologically) much like a normal human being. But as soon as we admit that Caligula doesn’t look much like a normal person, are our intuitions nearly as strong? We can test this question by supposing that there is an alien race of beings who look just like humans, who are made up of the same stuff, but who have ideally coherent, sadistic psychologies. Harman, for instance, asks us to imagine the case in which

[i]ntelligent beings from outer space land on Earth, beings without the slightest concern for human life and happiness. That a certain course of action on their part might injure one of us means nothing to them…In such a case it would be odd to say that nevertheless the beings ought to avoid injuring …Of course we will want to resist them if they do such things and we will make negative judgments about them; but we will judge that they are dreadful enemies to be repelled and even destroyed, not that they should not act as they do (1975, 5).

If Harman’s suggestion is correct here, then when a being has sufficiently different concerns from us, we stop trying to get a normative grip on them, and instead simply defend ourselves against them.

Sharon Street uses a similar strategy to argue that all Ideally Coherent Eccentrics (ICEs, as she calls them – Caligula, the anorexic, the blade of grass counter, future Tuesday indifference man) can be handled the same way (2009). The problem, according to Street, is that when these cases are offered in objection to an attitude-dependent theory of normativity, they are not given nearly enough consideration. Just as Caligula is not a normal person with a few tweaks, neither
is the ideally coherent anorexic. If any person were really so psychologically centered on a bizarre or scary project, then they may well seem more alien than human. Street also takes the alien analogy to be helpful, writing:

If a group of intelligent, ideally coherent aliens descended upon us and began trying to kill us for food or torture us for sport, would we feel intuitively convinced that they were making a mistake about the normative facts? Would we be more inclined to say “They shouldn’t be doing this” or rather just “How can we stop them?” When it comes to real-life people, those of us deeply committed to morality should always err in the direction of assuming that an alleged amoralist is making a mistake, for in real life such a person is in all likelihood inconsistent, wrong about many of the non-normative facts, and self-deceived. Most real-life human beings have moral feelings that can be tapped into, and in talking to them and others about their normative reasons we can appeal to that. The more we start to suspect that a person hasn’t even the tiniest trace of moral feeling, however, the more the alien species analogy will start to seem the right one (2009, 16)

Surely, not everyone will be convinced by the alien analogy. The real strength of the Harman/Street response to ICEs, however, is to point out how strange such people are. They aren’t mostly like us, with just a few differences; they are very different, and it’s not at all clear that the way we react to people in the actual world tells us how we should react to them. And none of this requires abandoning the lessons learned in the Hitler case – there is still much that we can say truthfully about Caligula, since it seems true, consistent with HCA, that Caligula is a monster and that something ought to be done about him. The appropriate response to Caligula, then, is to defend ourselves against him, punish him if we can, and tell one another what an
awful creature he is. On my view, we haven’t lost anything in this exchange, despite the fact that there is one thing that we cannot say: we cannot say that Caligula had reason to treat us differently. If that still seems a cost, it is a cost that I think the proponent of HCA can live with.

6.4 Defending Ambition: The Prevalence of Moral Reasons

At the beginning of this chapter, I noted that desire- or interest-based explanations of morality face a dilemma: any account that fails to justify the universality of normativity will be seen as unsatisfying, while any account that attempts to generate universal normativity from such humble beginnings will be the subject of incredulity. HCA, by holding that very many people have weighty or most reason to act rightly, manages to invite attacks from both sides of the divide. In the previous three sections, I have responded to the first worry. Although I almost certainly did not convince the committed realist or other objectists, I have explained why I and others (notably Gil Harman and Sharon Street) do not take the fact that a view is non-universalist to be a decisive criticism of it. In what follows, then, I turn to the second horn of the dilemma, and will attempt to argue that, though HCA does claim to explain a wide-spread reason to be moral, one should not find such a claim incredible.

Recall that HCA is simply the view constituted by marrying Scanlon’s criterion of wrongness, Schroeder’s analysis of what it is to be a reason, and a proportionalist theory of weighting. As a schematic view, there is nothing to find incredible; HCA simply states that morally permissible actions are justifiable ones, and that one’s reasons (and the weights of those reasons) are determined by one’s particular desires (and the strength of those desires). According to such a view, whether or not one has a reason to act morally depends on whether or not one has a desire that is promoted by so acting. Now, even as a schematic view, we might
suppose (as Scanlon did in (1982)), that on such a picture, many of us would have reason to be moral. Why? Because we desire to be able to justify ourselves to others. Or, if we want to include some less morally-sensitive folks, we might also point to more ‘Hobbesian’ sorts of reasons, like those grounded in the desire to not be thrown in prison or killed.

However, as I noted in Chapter 2, I would not yet find HCA so compelling if left in this, schematic, form. What led me to judge that HCA is genuinely compelling is the vindication of The Motivational Basis of Morality, which holds (in part) that some agent has a reason to be moral if and because she desires to have intimate relationships. That is to say: according to HCA, anyone who desires to have intimate relationships has a reason to be moral. And further, given proportionalism, anyone who has the desire for intimate relationships as among her strongest desires will have a weighty – often the weightiest – reason to be moral. To say this is not to deny that there are other reasons to be moral. In particular, the desire to act or to live justifiably still grounds a reason – often weighty for many of us – and so the addition of the intimacy-based reason simply adds to the explanatory power of HCA. While a desire to be able to justify oneself explains one kind of reason to act rightly – the kind that most (likely all) who would read this dissertation have and to which we reliably respond – the desire for intimacy explains a different reason. This less morally-relevant desire explains a reason that can ‘get a normative grip’ on a much broader group of people. In the language of Chapter 1, it helps us to navigate Prichard’s Dilemma, by showing how ‘the moral reason’ is actually complex, consisting of multiple reasons.

Although I will not rehearse the entire argument here, it will be helpful in evaluating the credibility of my claim (that very many people have weighty, or most, reason to act morally) to
quickly recapitulate the argument for The Motivational Basis of Morality. Although the argument in the dissertation took off from an example of Scanlon’s, the general intuition is an Aristotelian one, as it posits a connection between the ability to have important relationships and one’s moral commitment. So, while I admitted that both Aristotle’s and Scanlon’s particular claims were implausibly strong, I believe they point us towards something true. Aristotle famously wrote that the highest level of friendship could be had only between the virtuous (1999, Books VIII-IX), and Scanlon claimed that one could not be ‘true friends’ with someone who was uncommitted to justifiability (1998, 164-165). My question was why they would think these things – that is, what is it that grounds whatever plausibility such claims have?

Scanlon’s answer is that allowing ourselves to be so close to someone who saw no need to justify his actions would be ‘unnerving’. Thus, if a friend told me that he would steal a kidney on my behalf, I should be disturbed; sure, he wouldn’t steal my kidney, but that’s because he happens (right now!) to like me.

My interpretation of this insight is that a certain level of closeness to another person brings with it vulnerability, and so we tend to want assurance that the other person is committed to not acting in ways that would hurt us. In my language, then, it is not friendship that is special, but rather it is intimacy: whether it be with an intimate friend or a lover, exposing ourselves physically and emotionally provides the other with lots of opportunities to hurt us. It is thus reasonable that I be wary of so exposing myself to someone who is uncommitted to acting justifiably.

Now, if something like Scanlon’s or Aristotle’s claims were true, we would end up with a very strong conclusion: it would be the case that one could not have real friends (or, if my
intimacy amendment is correct, on could not have intimate relationships) if one was uncommitted to justifiability. As I admitted in Chapters 1 and 2, this seems implausibly strong. However, the intimacy/vulnerability connection seems real, which means that people will predictably be less likely to become intimate with someone who is uncommitted to justifiability. Others will see it as a risk, and so avoid it if they can. Certainly, we can drink wine with bad people, and even have casual sexual relationships with them. We can be close in many ways, and some people will marry or attempt to be best friends with vicious people. But the morality/intimacy connection suggests that many such unions will end badly, and so very many of us will try to avoid such relationships. What this means is that failing to develop a commitment to justifiability is a fairly reliable block to something that most of us intensely desire – namely, intimacy. And so, on a suitably liberal interpretation of the ‘promotion’ relation,12 becoming a justifier promotes the good of having intimate relationships.

My view of the relationship between intimacy and morality is much weaker than Scanlon’s or Aristotle’s, but it is strong enough to generate a surprising conclusion: in all those cases (which I have suggested are quite a lot) where becoming a justifier promotes the good of intimacy, the agent in question has reason (likely weighty) to ‘be a justifier’. In the language of Chapter 5, then, for a surprising number of people, the best strategy for satisfying one’s strongest desires will be simply to adopt a commitment to justifiability.

Weakened from Scanlon’s and Aristotle’s inspiration though it is, the argument above still relies on the belief that a lack of commitment to justifiability is at least a reliable (although

---

12 As discussed in 1.7.
likely not perfect) block to intimacy. And it is this claim that some will still see as too strong, and so which will draw some incredulity. As a result, the supposed connection between intimacy and morality will be challenged by counterexample. Now, since I abandoned the strong necessity claims of Aristotle and Scanlon, a single counterexample won’t work – one case where an agent was able to form intimate relationships despite being uncommitted to justifiability does not threaten the argument. However, if one could identify a sufficiently large number of cases in which an imperfect commitment to justifiability did not seem to serve as any kind of block to intimacy, then it would look as if my argument had made a mistake somewhere. In the following section, then, I will investigate the most common counterexamples offered to the intimacy/morality connection. My response to these examples will take me much of the remainder of the chapter, and will ultimately lead me to a discussion of other ways in which HCA seems unsatisfying.

6.5 Loyalty-Clan Counterexamples

For one who believes that there is a connection between a commitment to justifiable action and the ability to have intimate relationships, the most disconcerting cases are what I will call ‘Loyalty-Clan Counterexamples’, or LCCs.\(^\text{13}\) In what follows, I will provide a few different examples of LCCs, as the subtle differences between particular examples will affect how I respond. But in general, the features of an LCC are standard, as they posit a case where, by virtue of being part of a group or ‘clan’, individuals seem able to act in terribly unjustifiable

\(^{13}\) Thanks to Justin Weinberg both for the name, and for the first of very many instances of this objection.
ways towards outsiders while maintaining genuinely intimate relationships with other clan members. The most common cases that have been offered to me are as follows:

The Mafioso: If mobster movies and novels are to be believed, the mafia employs many members who are ‘enforcers’, whose job it is to do terrible things to non-family members when so ordered. The same Mafioso who is out extorting, bludgeoning and murdering during the day, however, is often depicted as having intensely deep friendships and a lover back home.

And since the Nazis are never far from a moral conversation, we can also add

The Nazi Family Man: Thousands of German soldiers were required to do terrible things on behalf of the Third Reich, including participating in the systematic murder of Jews and other minorities. Yet it seems plausible that many of these men worked hard at their killing during the day, and then went home to their families at night, where they enjoyed genuinely intimate relationships with their spouses.

In my responses below, I will use one tactic in order to respond to these first two cases. A lesson drawn from the Nazi case, however, will show that there are other LCCs that must be addressed in a different way. The final two examples highlight what I take to be the more troubling lesson of LCCs.

The Common Slave-owner: For hundreds of years in the American south, white Americans owned slaves and often treated them in terrible ways. Even some American heroes, such as George Washington and Abraham Lincoln, owned slaves. Presumably, however, very many slave-owners had intimate friendships and romantic lovers. Likely, even some of the violent and terrible owners had such relationships.
And finally:

**The Fierce Nationalist:** Many citizens of various states believe that whether another person is a fellow citizen matters morally. This kind of nationalism exists on a spectrum, from the moderate view that fellow citizens ought to be given some (even if minimal) priority in the distribution of resources, all the way to the view that a fellow citizen has a full suite of rights and privileges, while foreigners have little or no such standing. In a sufficiently large and like-minded community of those at the more extreme end – what we can call *fierce* nationalism – it seems plausible that one could reliably have intimate relationships with other fierce nationalists while engaging in racist or xenophobic behavior.

If these four cases (and we could likely generate many others) are something like accurate depictions of various lifestyles (Mafiosi, Nazi soldiers, slave-owners, and nationalists), then they raise a serious problem for HCA. In what sense could a lack of commitment to justifiability be a reliable block to intimacy if people in all of the cases above have no trouble forming intimate relationships while doing what they do?

In response, I have to begin by admitting something fairly depressing: that the desire for intimacy is sometimes best promoted by acting justifiably *to some group*, rather than by acting justifiably *simpliciter*.¹⁴ Now, there will be much to say about this admission. Ultimately, it will

---

¹⁴ Recent work in moral psychology provides an interesting context for this observation. According to Josh Greene (2013), we have good reason to believe that morality evolved to deal with a world of ‘tribes’, rather than to deal with a globalized community. On his view, then, our ‘point and shoot morality’ (Greene’s language) – that is, our instinctive moral reactions – commonly generate ‘loyalty-clan’ judgments (my language). Although I clearly cannot fully explore the relationship between HCA and a theory such as Greene’s, the kind of evolution of
explain why I must accept some cases in which a desire for intimacy does not provide an avenue for moral reasons to get a grip on an agent. That is unfortunate. However, this admission will not force me to a radical moral relativism (as it might seem), nor will it entail that I must uncritically accept any LCC. In the remainder of this section, I will focus on what I can say about some LCCs, and then at what point HCA must admit some counter-intuitive implications.

The general admission from the paragraph above is due to the fact that the intimacy/justifiability connection – if I’m right that it exists – is an empirical one. As described here and in Chapters 1 and 2, the connection exists because, as a matter of contingent fact, most of us want an assurance of good will prior to exposing ourselves to one another. In essence, the argument for such a connection works because people need to feel protected when they make themselves vulnerable. Through a commitment to justifiability, an agent is able to provide that security, and so such a commitment promotes the having of intimate relationships.

The problem brought to light by LCCs, then, is that there is another commitment that is able to make others feel secure and so that would promote the having of intimate relationships: namely, loyalty. For present purposes, we shouldn’t think of loyalty just as a pledge of allegiance to members of a group; but rather, a commitment to act justifiably to members of one’s group. As a mobster, I’m not allowed to extort members of the family, and as a fierce morality that Greene posits may explain one more virtue of a view like HCA: perhaps, previous to the full, rich morality that we now attempt to explain, there was a ‘primitive’ or ‘proto-morality’ that developed in the way Greene suggests. HCA, then, is able to explain both why we often have reason to act morally, and why those with a proto-morality often had reason to act proto-morally. Now, such speculation certainly does not help HCA to respond to LCCs now, but it does suggest that HCA has the theoretical virtue of explaining, with one theory, both moral reasons and ‘proto-moral reasons’. Thanks to Henry Richardson for helpful discussion on this point.
nationalist, I’m not allowed to discriminate against fellow citizens. But if you’re not one of ‘us’, then you’re fair game. Loyalty does not require that one abandon a commitment to justifiability; it requires that one restrict the scope of the class to whom justifiability is owed.

Within a loyalty clan, then, a commitment to loyalty will provide the same guarantee (to another member of the clan) of safety, and so a desire for intimate relationships will generate a reason to act _loyally_ rather than a reason to act justifiably. 15 Although I think that such an explanation does show that some LCCs work as criticisms of HCA, an evaluation of exactly how harmful we should see this criticism requires a closer look at the different kinds of cases.

The key point that I want to make in evaluating the LCCs is that, if loyalty is to provide a sufficient sense of security to allow for and promote intimacy among its members, then the restriction of justifiability to members of the group must make at least some intuitive sense to them. That is: Scanlon’s precise worry was that ‘warm feelings’ by someone are insufficient to make an intimate feel secure; we want, he suggests, to see that the other has a commitment to acting towards me in ways that I and others could accept. So the fact that I’m part of a group can’t make me feel secure if it amounts only to the other members liking me more. And indeed, the plausibility of LCCs is due precisely to their use of clans where membership of a group is of genuine importance. ‘Outsiders’ are _different_ from us. Different how? Well, if they aren’t owed

---

15 One might think that the ‘rather than’ claim is too strong: surely a desire for intimacy generates a reason _both_ to be loyal and to be a justifier, since they are two different ways of promoting the desired end. However, we can actually imagine that in some of these cases, being a justifier would undermine one’s ability to relate to other members of the clan, as the group norms explicitly forbid giving outsiders the same standing as one gives to insiders. And so for at least some of these cases, it may well be that a desire for intimacy is promoted by being loyal _rather than_ being a justifier.
justifiability, they must be *less than us*. In such groups, we would expect to find dehumanization of the other, and an irrational, emotional bias against outsiders. And this is largely, I think, what we find. Slavery provides what is perhaps the perfect example, especially due to its familiarity. Africans brought to America were literally thought to be sub-human. Of course they weren’t owed justification, because they were more animal than person.

What about cases in which the division between us and them is less stark, though? My suggestion is that, in such cases, we should expect the division to make less intuitive sense to its members, and so a commitment to loyalty would likely be less effective (and the group may well be unstable as a result). Consider again *The Mafioso*. Now it’s likely that different mob organizations work differently, but we can imagine a version of the case in which the enforcer was regularly called upon to eliminate members of the ‘family’ who had done something offensive. In addition, individuals who are not biologically part of the family can be married in, as well as ‘adopted’, if the organization’s leadership decides it prudent. In this case, the line between ‘us’ and ‘them’ is very fuzzy. In fact, it would be obvious to anyone that there is no *real* difference between a member of the group and an outsider. If I am in a relationship with the enforcer, I might be glad that he currently has no reason to kill me, and we might enjoy one another’s company. But none of the group’s members could really believe that everyone else is ‘less than human’, or something like that. Whether or not someone is afforded the status of ‘to be protected’ or ‘to be killed’ is a convention established by the organization’s leadership. And knowing this, what I know about the enforcer is that, however much he likes me, if ordered by the boss to but a bullet through my head, he would do it.
In this version of *The Mafioso* case, I no longer find it counter-intuitive to believe that the mob enforcer may have trouble forming genuinely intimate relationships, or that, even if others do get close to him, perhaps they had good reason not to. Loyalty seems most able to replace a commitment to justifiability when members of some loyalty clan are able to easily and intuitively believe that their inclusion, and others’ exclusion, from the group is principled.

In some ways, the case of *The Nazi Family Man* is thus the most interesting, as Nazi propaganda is good evidence that Hitler understood the connection between dehumanizing the other and promoting loyalty. However, even despite these efforts, many stories emerged from Nazi Germany about the difficulty of training people to exterminate another people. Indeed, this is likely part of why the ‘gas showers’ proved so successful as a means for carrying out the Party’s evil project: instead of looking at a Jew while shooting them, soldiers were able to just march them out of sight and then gas them to death. And the difficulty of treating Jews and other minorities like garbage is understandable precisely because, despite Hitler’s propaganda, it is difficult to believe that entire races of people actually aren’t people. When the naming of some group immediately made them the target of the Third Reich, it was likely hard to believe that there was any real difference there. And so I would predict that the rise of the Nazi party in Germany led to many devastated relationships, many broken families, and many ruined people, even among the perpetrators of violence, as it was likely difficult for at least some soldiers to convince themselves that their actions were justified. And that hypothesis doesn’t sound counter-intuitive at all. The case of *The Nazi Family Man* is thus a complex one. There likely were families that were able to convince themselves thoroughly of Hitler’s message, and so a commitment to loyalty worked to preserve intimate relationships. However, we can also guess
that there were many cases of broken Nazi men, and shattered Nazi relationships, and that is
evidence that a complete lack of commitment to justifiability may well devastate one’s
relationships, even in extreme cases like that of Nazi Germany.

At this point, I have done what I can to push back on the LCCs. In at least the first two, I
do not think the case is so clear-cut. Some loyalty clans may work some of the time in the world
of mob bosses and tyrant regimes, but the very argument given on behalf of the
intimacy/justifiability connection predicts what seems plausible, that many of these cases will
end badly for the people involved.

However, this leaves the final two cases of The Fierce Nationalist and The Common
Slave-owner, both of which seem plausibly to go through. That is: it seems to me that in each of
these cases, a desire for intimacy generates a reason to be loyal rather than to be a justifier, and
so the arguments of Chapters 1 and 2 have an even further restricted scope. In what follows, I
will show that this admission is just a slight extension of what I have already admitted – that the
view fails to vindicate the universal normativity of morality. It does not, as it might appear,
imply that HCA is a version of moral relativism, nor does it rob us of our ability to condemn the
bad practices of other cultures.

6.6 Why HCA isn’t (Morally) Relativist

It looks as though nationalists and slave-owners may have weighty reason to act loyally
rather than justifiably, if they strongly desire intimacy. A common reaction to this admission is
that it makes HCA sound culturally relativist,\(^{16}\) in that it carves out groups of people – at least

\(^{16}\) I am here specifying *cultural* relativism in order to clearly distinguish between the view under
consideration – a first-order, moral theory – from what we might call ‘metanormative

238
sometimes according to their culture (‘tribes’ such as fierce nationalists and Southern slaveowners) – and admits that what they have most reason to do is different from what the rest of us have most reason to do. Further, then, since my opponent likely believes that cultural relativism is false, this implication is a reductio of HCA.

There are several things to say in response. The most pressing is that, despite the entrance of ‘groups’ or even ‘cultures’ into the discussion, the admission concerning LCCs does not entail that HCA is a culturally relativist view. Cultural relativism is a form of moral relativism, which holds that whether an action is right or wrong depends on the accepted practices of the actor’s culture. If HCA was a culturally relativist view, then, it would say about nationalists that discrimination is permissible, or even obligatory, because that is what one does in that group. However, HCA does not say this. HCA says that an act is impermissible if it is unjustifiable; discrimination is unjustifiable, and so it is impermissible. The conclusion that LCCs force HCA to is that a nationalist may have most reason to discriminate against a non-citizen, and so most reason to act impermissibly.

It may well be that one is disconcerted by this separation of reasons from morality, but I have already admitted in the first sections of this chapter that HCA is ‘externalist’ in this way. The fact that an act is wrong does not always provide a reason not to do it, and the case of LCCs has simply expanded the implications of this claim from individuals to groups. The expansion is relativism’. HCA is clearly relativist metanormatively, as it holds that whether or not one has a reason is relative to her particular desires. As I hope becomes obvious throughout this section, a view can (and HCA is) metanormatively relativist while being ‘culturally’ or ‘morally’ non-relativist. In such a case, whether or not an agent has a reason will be relative in some way, but the content of morality will not be relative to the accepted norms of a community.
important, however, as it forces HCA to admit that it is not only unrepentant psychopaths, moral monsters, and the perfectly evil characters of our imagination who may have weighty reason to act wrongly; in fact, it may be entire cultures of well-meaning but mistaken people. After all, LCCs challenge that many American heroes did not have an intimacy-grounded reason not to own slaves, since their desires for intimacy would have grounded reasons to be loyal rather than to be justifiers. And this claim does seem (at least at first) to be more costly to HCA than the admission that the Ideally Coherent Caligula, if he were to exist, would have no reason to be moral.

Notice, however, that LCCs do not force me to admit that many good people had no reason to be moral; but rather, only that they had no intimacy-based reason to be moral. George Washington, for instance, may not have had an intimacy-based reason to free his slaves for at least part of his life. Because of the norms of the day, his having slaves would not have impeded his ability to have intimate relationships, as slaves were ‘outsiders’ and taken to be less than fully persons. However, there seems to be good historical reason to believe that George Washington had an exceptionally strong desire to live a justifiable life, and if so, then he had weighty reason to act justifiably grounded in this immediate desire. And in fact, Washington himself seems to have been aware of this reason and to have struggled with it, as he was an early advocate for the abolition of slavery, and he declared in his will that his slaves be freed upon the death of his wife.\textsuperscript{17} Thus, even in an environment that allowed for and promoted loyalty over justifiability, Washington recognized the reason not to own other human beings.

\textsuperscript{17} Washington’s Last Will and Testament can be found in Vol 37 of (Washington and Fitzpatrick 1745-1799, 276-277).
A virtue of the ‘complex theory of moral motivation’ defended in this dissertation, however, is that it does an excellent job of accounting for moral reasons generated by *either* the morally relevant desire for justifiability, or the less morally relevant desire for intimacy. The latter allows us to explain why many who lack central moral desires nonetheless have reason to act justifiably, while the former allows us to explain how those who do have central moral desires have reason to act justifiably, *even when the norms of their cultures would allow them to live well without doing so*. This leaves the group of people who do not seem to have any reason to act justifiably much smaller than it might appear from initial consideration of LCCs. The Nazi family men, fierce nationalists, and slave-owners still have reason to act justifiably, so long as they desire to live justifiably; and so long as we think that many of them were not so different from us, then many of them did have such desires.

Another aspect of the worry that HCA seems relativist is that it may seem that we are left unable to condemn morally bad norms or cultures as a whole. But one thing that we can see immediately, now, is that because HCA isn’t in fact culturally relativist, we *can* say that certain practices and norms are wrong. Slavery is wrong because it is unjustifiable. Further, then, we should recall all of the judgments that HCA can make about individuals who lack the ‘moral desires’, as they transfer over to the current contexts. The norms of some cultures are bad, and the actions of vicious slave-owners are monstrous. Further, America should have abolished slavery, and it is good that it did. Abolishing slavery was an instance of moral progress.

6.7 Conclusion

This final, substantive chapter is long, and attempts to make at least some headway against a fairly wide variety of concerns and objections. If HCA is to be taken seriously as a
primary candidate view of moral reasons, it must be shown to be less costly than it might otherwise seem. And while I certainly do not believe that I have offered decisive responses to each, I do believe that what I have said here provides reason to be: (1) less concerned about the non-universality of HCA; and (2), more optimistic that the argument from intimacy is not implausibly strong. In mounting this defense of HCA, I imagine I may have frustrated some of my interlocutors, who may wonder whether I have argued myself into a very strange position. In the following, concluding chapter, I will use this worry as a springboard for my final reflections on HCA, as addressing the worry will bring the discussion full circle, back to some of the insights with which this investigation began.
CONCLUSION

Having articulated and defended a Humean Contractualist Account of Moral Reasons, I would like to close this dissertation with two, final reflections on the accomplishments of the view. The first of these concerns a suspicion one might have that my responses to criticisms in the previous chapter rendered the overall view very strange. In particular, I will explain how the ‘moral externalism’ of HCA is still able to explain the special negative status accorded to wrong actions. Secondly, I will look back at the total view of moral reasons developed here, and re-raise the issue of HCA’s standing as a ‘backup’ theory of moral reasons. While I think that the articulation of the view in its more modest form – as a view that should be comforting, even to those who find it unsatisfying – is helpful, I also believe that a complete look at the costs and benefits of the account should lead us to consider it as a primary candidate view.

7.1 Morality and Justifiability

At the conclusion of Chapter 6, my opponent is likely to be frustrated. It seems, one might think, that HCA should be forced to accept more costs. There is something sneaky about the way that I have been able to avoid what seem like difficult questions for one who attempts a project as philosophically modest as mine. And while I hope that accepting fewer costs than expected is a virtue of my view, I understand the frustration.

The discomfort with my answers – for those who experience it – likely comes from the fact that I began this investigation with a large assumption: that moral action is justifiable action. And while I have gone to great lengths to argue for the view that we can explain ‘the reason to be moral’ utilizing philosophically modest assumptions, one may now be wondering what I have gained by doing so. After all, I have claimed that unjustifiable acts are still wrong, even when
one has no reason not to do them, and so one might wonder what it adds to call an action wrong in this case.

Interestingly, this concern brings us full circle, as it can be seen as a version of the ‘redundancy’ objection first mentioned in Chapter 2. There, we saw that many philosophers object to Scanlon’s contractualism on the grounds that it is either viciously circular (presupposing the moral content needed to generate moral conclusions), or redundant (as the contractualist machinery is an unnecessary addition to ordinary moral reasoning). As a reminder: these two criticisms are related in that they challenge the ‘two-level’ nature of Scanlon’s view, since he requires normative reasons as input into a contractualist machinery that is tasked with generating permissibility criteria. The circularity objection is easily handled, I argued, as it is clear that Scanlon need not import obligation or permission, and so (to borrow Street’s helpful terminology), the materials of construction are not identical to the results of construction (Street 2008, 215). Contractualism is thus not circular because the contractualist machinery doesn’t presuppose moral rightness or wrongness – it is precisely what is required in order to turn reasons into obligation, permission and prohibition.

The redundancy objection, though, is slightly more challenging. It asks why we need the contractualist’s machinery, when there already exist reasons that seem to do the kind of work that we need morality to do. In essence, it asks contractualism: what is gained by the addition of the particular reason to act justifiably? After all, if it’s justifiable, then there must be reasons in favor of doing it (or, on Scanlon’s formulation, reasons to reject principles that would disallow it), so why don’t we just rely on those reasons?
HCA does not hold that everyone always has a reason to act justifiably, so it might look as though it escapes a redundancy objection. However, a very sharp series of articles by Philip Stratton-Lake and David McNaughton and Piers Rawling show that a version of the redundancy objection may apply to Scanlon’s contractualism, even if it does not claim that wrongness is reason-giving.\(^1\) Stratton-Lake begins by arguing that Scanlon can avoid the redundancy charge by claiming that wrongness, like value (on Scanlon’s account), is a buck-passing notion – that is, that it does not provide a new reason, but rather it serves as an indicator that other reasons are present.\(^2\) Now, HCA is not a buck-passing account of wrongness, but it shares with buck-passing accounts the feature that (at least sometimes, according to HCA) an act can be wrong without its wrongness providing a new reason not to so act, and that this status as wrong is explained by its relation to other reasons (namely, the reasons that make some act justifiable or not).

McNaughton and Rawling, in response to Stratton-Lake, argue that even if Scanlon adopted a buck-passing account of wrongness, he would still be open to the redundancy objection (2003). Why? Not because a new, moral reason is redundant given some already existing reasons; but rather, because the mere fact that an act is unjustifiable – which is to say

---

1 See Stratton-Lake (2003a), McNaughton and Rawling (2003), and Stratton-Lake (2003b).
2 As I showed in Chapter 2, I do not think that Scanlon must make any changes to his view in order to evade the redundancy objection, since I think it fails to note the genuine addition of distinct, interpersonal obligation onto the scene. And in fact, the comments on which I will focus by Stratton-Lake in his reply to McNaughton and Rawling precisely show the kind of move that I think Scanlon could have made without moving to a buck-passing account. But this doesn’t actually matter for present purposes. What is helpful about the dialogue being discussed is Stratton-Lake’s insight into what is added by a contractualist notion of wrongness, even if it is not independently reason-giving.
that it’s wrong – is redundant. Reporting such a fact, say McNaughton and Rawling, doesn’t add any new information; they argue that

to say that the cruelty of torturing others for fun provides us with a reason not to reject principles that forbid such torture does not seem to add anything to simply saying that its cruelty provides us with a reason not to engage in it. To say that phi-ing is forbidden by principles that cannot be reasonably rejected is supposed to add something to the thought that there is sufficient moral reason not to phi. But it does not seem to…. [T]o say that there is sufficient reason not to reject principles that forbid phi-ing is a prolix way of saying that there is sufficient reason not to phi, and the reasons are the same in both cases. (2003)

Indeed, although McNaughton and Rawling do not suggest it, one might think the case is even worse for contractualism as a buck-passing account. If some act’s being unjustifiable does not even provide a reason not to phi, then it cannot even be argued that the reason is a non-redundant one. Against Stratton-Lake, then, we might think that a buck-passing contractualism is in even more trouble with redundancy than Scanlon’s own contractualism, which is exactly why the objection to HCA arose at this point: if wrongness doesn’t provide reasons, then it is unclear that it does anything useful at all. To call an act wrong, says the challenger, is redundant in terms of moral reasoning, and unhelpful in terms of understanding why we should care about an act’s unjustifiability.

I don’t think that calling an act wrong, in the contractualist context, is redundant or unhelpful, even if wrongness isn’t reason-giving in the circumstances. In the first place, think back to the responses I gave to the Hitler and Caligula cases of Chapter 6, and we can see that
similar claims apply here. When people act wrongly, they are (at least sometimes) acting in ways that provide the rest of us with reasons – reasons to resist the wrongdoer, to interfere with his plans, or perhaps to conclude that he has forfeited certain rights normally accorded to him, such as the right against the person who violently tries to interfere with his plans. An act’s being wrong, even if it doesn’t provide the acting agent with a reason not to do it, clearly has normative impact on the rest of us, affecting our reasons and perhaps even the rights and protections normally accorded to the actor.

However, my challenger may rightly point out that it’s not clear that these implications are due to the wrongness of the act, per se, but rather to contingent features of various wrong acts. What we might expect from an account of wrongness is an account of the special normative implications of wrongness – why is it that we would be given reasons in relation to wrong acts, rather than merely harmful, obnoxious, rude or otherwise inappropriate acts? An account that explained how wrongness always provided overriding reason not to act would discharge this burden, but HCA doesn’t do this. And this is where I think appeal to Stratton-Lake’s observations is particularly helpful, as his response to McNaughton and Rawling explains precisely what is special about wrongness. He reminds us that one of Scanlon’s goals was to explain the importance and priority of morality, and that this isn’t accomplished merely by adding a reason – it’s accomplished by identifying the moral with a substantive good, about which each of us has reason to care. Moral concern, on the contractualist’s formulation, is a concern to justify ourselves to others; and, as Stratton-Lake puts it, “A concern to justify

---

3 See further my discussion of importance and priority in Chapter 1.
ourselves to others on grounds they could not reasonably reject is a proper way of valuing rational, self-governing agents, and is thus a way of respecting their value as rational autonomous agents” (2003, 336). That is: on a buck-passing version of contractualism, the fact that some act is unjustifiable does add something to the already present facts concerning what reasons there are not to do it; it ‘characterizes’ the reasons already present in a way that “makes possible a Kantian explanation of their importance and priority” (2003, 336-337).

Pamela Hieronymi picks up on this thread of reasoning and argues powerfully that Scanlon’s Kantian link between moral motivation and moral content is precisely one of the best reasons to believe the contractualist view. The redundancy objection isn’t just mistaken, on her reading of Scanlon; it gets things almost exactly backwards. Not only does contractualism have an explanation of what wrongness adds to moral reasoning, but this explanation is one of the most attractive parts of the view. As Hieronymi puts it,

by acting wrongly, you have acted in a way that not only neglects the interests of those you have wronged, but that also denies their standing to (partly) determine the terms on which we each shall live. You have, thereby, acted in a way that fails to accord them a certain form of respect…You have, as Scanlon sometimes puts it, violated the terms of a relationship of mutual regard, the terms on which a kind of mutual recognition is possible, and so you have put yourself in a very different relation to your fellows. (2011, 108; emphasis hers)

Whether morality is reason-giving or not, then, according to Stratton-Lake and Hieronymi, Scanlon’s Kantian view of moral motivation provides a powerful explanation of what it adds to say that some act is unjustifiable, and so, wrong.
In a dissertation that has attempted to modify Scanlon’s view in a ‘Humean’ direction, it might be surprising to find me appealing to a Kantian solution. However, as I hope will become clear, the very insights with which I began this project – Scanlon’s moral motivational account – point us towards a promising Kantian story about what it adds to call something ‘moral’. Thus, though this stance may in fact be surprising, I do not see any tension between a Humean metanormative story and a Kantian first-order moral theory. In fact, the thesis that they can be fruitfully joined is precisely the contention of this dissertation. It seems only fitting, then, that I close with a few thoughts on just how ‘Kantian’ HCA is.

Because I partially dodged the objections of the previous chapter by leaning on the ‘moral externalism’ of HCA, my exasperated interlocutor has demanded to know what I gain by saying that an agent’s act is ‘wrong’, even when the agent has no reason not to do it. My response, in line with Stratton-Lake and Hieronymi, is that we gain a plausible story about our concern for right action and about why we react the way we do towards immorality. Those who disregard the requirements of justifiability, whether or not they have reason to do so, have exhibited to the rest of us that they do not see us as deserving justification. They do not see that you, and I, and all others have standing to partially determine the grounds on which we all act. I may have reason not to listen to pop country music (and I do!), but when a friend or stranger wants to listen to country, the comment that such a person chooses bad music doesn’t matter all that much. It doesn’t, we might say, add anything new to my deliberation about what music ought to be played. But that’s because music doesn’t have special importance or priority in our lives. As Scanlon puts it: being blind to good music is different from being blind to the reasons one has to act justifiably. The latter blindness “makes a more fundamental difference because
what is in question is not a shared appreciation of some external value but rather the person’s attitude toward us – specifically a failure to see why the justifiability of his or her actions to us should be of any importance” (1998, 159).

The contractualist believes that moral action – at least in the central class of cases concerning what we owe to each other – is justifiable action. And in the same way that our particular, actual desires and concerns hook up to generate reasons to act justifiably, they hook up to generate reasons to care about others’ commitments to justifiability as well. Morality is important, because it is about the standing I am given as a person, and immorality is important because it reveals the standing that someone didn’t accord me. In the same way that many of us are likely to see living together on justifiable terms as a good worth pursuing, we are also likely to see it as a good worth defending when others threaten it. Even if Hitler didn’t have reason to act other than he did, it matters that what he did was wrong; he violated standards that we could all sign on to, and so told all people that they didn’t have standing to negotiate our shared terms of living. We should, and typically do, want to know when this happens, because Hitler wouldn’t make a good friend, neighbor, lover, or even acquaintance. And so Hitler’s violation of justifiability gives us reasons to change our relationship with him. And of course, if we respond properly to this reason, then, so long as he wants us as friends and neighbors, we will have given him a reason to act justifiably.

---

4 Drawing out this implication helps to make clear how nicely Scanlon’s later work on blame (2008) complements his contractualist view. Although he is careful never to rely on his own contractualist view while defending his theory of blame, the investigation here (and in Hieronymi (2011), which greatly influenced my thinking in this area) really draws out the way in which a relational theory of reactive attitudes can help to fill in the contractualist picture.
7.2 HCA as a Primary Candidate View?

Although I have, at multiple points, asked for patience in sketching HCA by reference to its value as a ‘backup’ theory of moral reasons, it is likely unsurprising that I take it to have more promise than that. The goal of utilizing the idea of a backup theory was to keep my opponents’ attention long enough to run through a genuine consideration of the virtues of HCA. As I observed in the Introduction, it seems to me that many philosophers reject moral theoretic accounts as soon as they come to what seems to be a counter-intuitive conclusion. But this model is hasty and unhelpful. And so an invitation to consider HCA as a backup view was intended to highlight the fact that one’s preferred account might, after all, be wrong, and so there is a real benefit to exploring how much one can do when starting from humble beginnings. However, now the investigation is closing, and the whole view is in sight, so I leave off the rhetorical strategy of asking whether HCA is a good backup theory, and ask now whether it is a plausible theory in its own right.

I have, throughout this dissertation, sometimes referred to the ‘costs and benefits’ of HCA, and to the concern of my opponents that the view’s costs are simply too high. I do not know what it would look like to run a formal cost-benefit-analysis on a philosophical theory, or to argue with any confidence that, given the benefits and costs of a particular theory, it ought to be accepted or rejected. Such an idea seems to imply more precision concerning what we should expect from a theory than is reasonable. However, it does seem potentially helpful to draw out the costs and benefits clearly and concisely for a particular view, in an effort to help the reader perform their own evaluation. As I will reveal below, I think that the overall modesty and
strength of HCA, combined with what seem to me to be non-decisive costs, render the view a promising candidate account of moral reasons.

Although different philosophers will weight such a consideration quite differently from one another, a major strength of HCA is its philosophical modesty; it has, as I sometimes put it, relatively low buy-in. Virtually anyone can accept at least formal contractualism (the view that justifiable action is permissible action, but which does not take a stand on the direction of explanation between the two), and that is all that is required in order to utilize HCA as a theory of moral reasons. The Humean Theory of Reasons, then, while surely in fact rejected by many philosophers, is among the most ontologically minimalist ways to explain the existence of normativity. So long as one accepts that there are facts and desires, one believes in everything needed to utilize the theory. Further, it seems overwhelmingly plausible that at least some reasons are explained in the way HTR explains them, so it is not a universally bad model of normative explanation. Finally, although I readily admit that my substantive arguments for other aspects of the view – for instance, that a commitment to justifiability promotes the desire for intimacy, or that Hobbes’ Fool can be answered – are non-obvious, I never abandoned the method of modesty. If there are cases in which a commitment to justifiability does not promote intimacy, then HCA does not generate that (particular) reason to act rightly in that case; and if a few people – talented con artists, say – are genuinely sufficiently different from me and my friends, then it may be that my solution to the Fool does not work for them. In all of these ways, HCA is a very easy view to believe, with no mystery surrounding any of the steps along the way.

Of course, no amount of philosophical modesty will render a view plausible if it fails to explain what it is supposed to explain. And that, I take it, is the real worry with HCA. But
before cataloguing the costs of the view, I want to note what it does do in the way of explanation. HCA, in my view, provides a novel and powerful ‘complex account of moral motivation’, which is uniquely well-suited to answer the challenge of what Scanlon calls ‘Prichard’s Dilemma’. By invoking both intimacy and the good of living with others on justifiable terms in its explanation of the reason to be moral, HCA is able to account for the way in which ‘the reason to be moral’ both explains why those who care about morality act as they do, as well as why (nearly) everyone has a reason to care about morality. HCA also provides a unique and satisfying way of accounting for both ‘Humean’ and ‘Kantian’ intuitions, by explaining how desire is normatively relevant, while still holding that there is an important role for the value of the rational agent. Finally, HCA does all of this, while generating the intuitively correct judgments about whether one has a reason to act rightly in very many normal cases, although the actual number of cases in which it ‘gets it right’ is largely an empirical matter.

This last claim, however, is where the action is, as my opponent will finally get to lodge her worry that matching our intuitions about moral reasons in ‘very many’ cases is simply not good enough. HCA is, after all, committed to the view that the Ideally Coherent Caligula has no reason not to torture, that Hitler may have had most reason to lead the Third Reich, and that at least some, vicious slave-owners had no reason not to own slaves. And although I admit that the view would be more satisfying if it did not yield these conclusions, at least some of my opponents will likely make the much stronger judgment that getting the correct conclusion in the cases of Caligula, Hitler and slave-owners is non-negotiable; that is, it is not a cost that we can choose to accept in a theory.
I simply don’t know if this last claim is true, or what could decisively ground such a judgment. And this leads me to end the dissertation on a rather unsatisfying, but honest, note. My arguments in Chapter 6 have shown that the number of cases in which HCA generates a counter-intuitive conclusion is limited in kind, but what they also show, I think, is that those cases are not the ones that have formed our intuitions about moral reasons. I do not hang out with fictional characters, Hitler, or vicious slave-owners; instead, my world is populated by people who have a strong desire to act justifiably and to live with others on terms that all could accept. Even more clear is that those people, familiarity with whom has informed my philosophical theorizing, strongly want intimacy in their lives. HCA predicts exactly what seems clearly true to me, which is that the kind of people I have always known, and who have molded my intuitions, have very weighty – likely most weighty – reason to act justifiably. This, I think, is an impressive conclusion.

I imagine the same prediction holds for anyone who would be exposed to the arguments in this dissertation: the people who populate your world will also, according to HCA, have weighty or most reason to act justifiably. These considerations lead me to believe that the virtues of HCA are substantial, and that the costs are not so great, nor are they clear. It may simply be that the cases where HCA seems to ‘get it wrong’ are cases about which our intuitions are less reliable. And so, without recourse to a cost-benefit-analysis for theories, I cannot do more than to say: that seems to me to make HCA a pretty good view, plausibly deserving of real consideration among other, competing theories of moral reasons.
BIBLIOGRAPHY


