Wanting to Do What's Right

by

Lyndal Jennifer Grant

B.A. Philosophy and Development Studies
University of Melbourne, 2007

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2023

Author _____

Department of Linguistics and Philosophy
August 2, 2021

Certified by _____

Kieran Setiya
Professor
Thesis Supervisor

Accepted by _____

Brad Skow
Laurance S. Rockefeller Professor
Chair of the Committee on Graduate Students

Wanting to Do What's Right

by

Lyndal Jennifer Grant

Submitted to the Department of Linguistics and Philosophy on
August 2, 2021, in partial fulfillment of the requirements for the
degree of Doctor of Philosophy in Philosophy

Abstract

It may seem obvious that good people want to do what is right. But moral philosophers disagree about whether it is virtuous to be motivated to do what is right *as such*. Some, inspired by Kant, argue that wanting to do what is right as such is always morally praiseworthy. Others claim that such a desire amounts to a kind of *moral fetishism*.

This dissertation lays out the groundwork for a new way of thinking about what it is to want to do what is right as such. The central task (which is the topic of chapter 1) is to provide a new account of moral fetishism that allows us to maintain what I take to be the natural view: it is not always wrong to want to do what is right as such (though it sometimes is). I argue that whether wanting to do what is right as such is virtuous or morally fetishistic depends on the deeper structure of the agent's motivations. What makes the fetishist a fetishist, I argue, is that they want to do what is right *whatever rightness might be*. By contrast, the good person's desire to do what is right is conditional on their substantive conception of right action being at least approximately correct. This account allows us to resolve seemingly conflicting intuitions about cases of wanting to do what is right, and also suggests a more general account of how the contents of our desires depend on our beliefs together with further features of our underlying motivational states.

Chapter 2 takes a deeper dive into the nature of desire contents, providing an independent, disposition-based argument for a thesis on which my account of moral fetishism depends: that two people can both want *p*, but in wanting *p*, nonetheless have desires with different contents. Chapter 3 then shows how my account of moral fetishism creates trouble for prominent theories of moral worth. The upshot is that any adequate account of moral worth will need to place additional constraints on the content of the desires that ultimately explain why the agent acts as she does.

Thesis Supervisor:	Kieran Setiya
Title:	Professor of Philosophy

Acknowledgements

Thanks go first and foremost to my committee chair, Kieran Setiya, for his guidance, support, and patience as I journeyed through many, many iterations of my view. Thanks also to my other committee members, Tamar Schapiro and Alex Byrne, for their feedback and advice, often on very short notice. Thanks to Augustin Rayo, Sally Haslanger, and Judy Thompson for their encouragement in my early days of graduate school; it is hard to overstate what a difference this made.

Graduate students in philosophy at MIT are a remarkable bunch. For their friendship and for countless helpful discussions, I'd particularly like to thank Arden Ali, Ekaterina Botchkina, Nilanjan Das, Brendan De Kenessy, Ryan Doody, Cosmo Grant, Sophie Horowitz (buddy!), Abby Jacques, Allison Koslow, Rose Lenehan, Matthew Mandelkern, Daniel Munoz, Sofia Ortiz-Hinojosa, Milo Phillips-Brown and Bernhard Salow.

And now to my family. Thanks to my parents, Ena and Alan, for being unfailingly supportive, even from so far away. And to my sisters, Kath and Rachel, for always finding a time to Skype and a way to visit. Thanks also to Martha Roberts, for her friendship and invaluable support over the last year. You have gone above-and-beyond the call of duty, and David and I are forever grateful.

Of course, huge thanks go my partner David Gray Grant, for his years of love, support (emotional and tech), and his enormously helpful feedback on so many drafts.

And to my daughter Mira, thank you for (usually) sleeping through the night. I love you more than you could know.

Introduction

We are not the ideally rational agents so commonly imagined in textbooks on moral theory. We often find ourselves unsure of what we morally ought to do, either because we don't know how to weigh a number of competing moral considerations, or because we don't know which moral considerations apply in a given case. Though we may aspire to the character of the *phronimos*, even the most morally sophisticated among us are sometimes *morally uncertain*. I take this much to be fairly uncontroversial.

What is more controversial is the further claim (which is central to my argument in chapter 1), that when someone doesn't know what they ought to do, there isn't necessarily anything *morally fetishistic* about their wanting to do what is right. I think that this claim too, is intuitive, though it conflicts with a prominent view which says that wanting to do what is right *as such*, or *de dicto* is invariably morally fetishistic.

The central question that I set out to answer in this dissertation is this: what is moral fetishism? I've already indicated at least part of my answer to that question: it is not wanting to do what is right as such. And yet those who hold this mistaken view have some compelling examples up their sleeve. Consider Bernard Williams' famous example of a man who saves his drowning wife, rather than a similarly placed stranger, "because it is my wife, and situations like this it is morally permissible to save one's wife." Or consider Michael Stocker's example of a friend who visits you in hospital not because she is concerned for you in particular, but because she believes it is her moral duty. In each of these cases, the agents' explicitly moral motives seem to detract from the goodness of their motivations, rather than reflecting well on them.

I don't think we should dismiss the significance of these cases. But nor do I think that we should, in trying to accommodate them, brush aside our commonsense reactions to cases of moral uncertainty. This leaves us with a puzzle: why do some cases of wanting to do what is right as such strike us as perfectly fine while others seem morally fetishistic?

In what follows, I argue that we do not need to choose between wholesale endorsement or condemnation of wanting to do what is right as such. Ultimately, given the strength of our intuitions about the aforementioned cases, I think that an account of moral fetishism (or moral worth, for that matter) that forces us to dismiss them does worse, *ceteris paribus,* than one which shows them to be justified.

But rather than rehearse the arguments to come, I want to say something about an underlying assumption that I suspect is at work in much of the extant literature on this question. At the risk of oversimplifying, we can see the philosophical literature as divided between those who claim (with Kant) that being moved by a desire to do what is right as such is *always* morally praiseworthy, and those who claim (motivated by cases like Williams' and Stocker's) that such a desire is *always* morally fetishistic. Despite their differences, these approaches share a commitment to treating all instances of wanting to do what is right as such in the same way.

I think of this as a kind of "common factor" assumption.[1] The assumption is that – whatever the differences between various cases of wanting to do what is right as such – in each case the desire to do what is right makes the same contribution to the moral status of the action. If it is morally fetishistic to want to what is right in one case (as in Williams' and Stocker's examples), then it is morally fetishistic in every case. If being moved by a desire to do what is right as such is not sufficient to confer moral worth in one case (as in Arpaly's case of the Nazi who wants to do what is right) then it is not sufficient in any case. Insofar as there are differences in our evaluation of these agents' motives, those differences must be attributable to features *other* than their desire to do what is right. After all, the desire that motivates the agent is the same in each case.

Each of the papers in this dissertation rejects this assumption in its own way. Rejecting it gives us a new way to think about wanting to do what is right as such. My account of moral fetishism in chapter 1 ("What's Wrong with Wanting to Do What's Right?") relies on the claim that although the moral fetishist and the good person may both want to do what is right as such, in wanting to do what is right, they want different things. And this means that (even setting aside their other differences) their desires to do what is right are *not* morally on a par. On my account, the reasons that explain why a person wants to do what is right make a difference to the satisfaction conditions of their desire—or to put the thought in less technical terms, to what it takes for them to get what they want. Chapter 2 ("Getting What You Want") provides an argument for the more general claim that two people can both want *p*, but in wanting *p* nonetheless have desires with different contents. The chapter therefore fills out the background view of desire contents that I draw on in setting out my account of moral fetishism in chapter 1. Finally, chapter 3 ("Moral Fetishism and Moral Worth") argues that rejecting the common factor assumption has important implications for how we should think about the nature of moral worth, in addition to moral fetishism.

I discuss the common factor assumption in only a few places in the papers that follow, and there only briefly. However, it (or rather, my resistance to it) is a central thread running through all three chapters. I suspect that something like the common factor assumption prevents us from seeing more satisfying solutions to a variety of problems concerning moral motivation in ethics, though I don't explore that broader set of problems and solutions here in anything but the broadest terms. I hope, at least, to have shown that rejecting the common factor assumption sheds new light on what we want when we want to do what is right.

[1] I borrow this terminology from discussions of disjunctivism about perceptual experiences.

Chapter 1

What's Wrong with Wanting to Do What's Right?

We often find ourselves unsure of what we morally ought to do. Sam wants to give her children the best start in life, but she isn't sure if the expense of sending them to private school is consistent with her obligations to help others in need. Xander wants the Democrats to win in his home state, but he knows his vote will make no difference to the outcome of the election, and so he isn't sure if he is morally obliged to vote anyway. Helen had always thought that everyone in her housing cooperative should get an equal say in whether new members are admitted, but now she suspects that some members are being influenced by racial bias, and she isn't sure if that means that their preferences should be dismissed.

In cases like these, it is often natural (and I will argue, morally unobjectionable) to want to do what is right, despite not knowing what that is. But this seemingly uncontroversial claim—that sometimes, good people want to do what is right *as such*—is at odds with an influential line of thought concerning "moral fetishism."

Perhaps the best way to get a sense of the phenomenon of moral fetishism is to consider some paradigmatic examples. Williams' (1981) case of a man who must choose between saving his drowning wife or a drowning stranger. The man saves his wife, and his motivating reason is: that it is my wife, *and in situations like this it is morally permissible to save one's wife*. He does what we are inclined to think he ought to have done, but something about his motivations seems amiss. According to Williams, he has "one thought too many": he is moved by the thought of the rightness of his actions *as such,* when he should be moved instead by direct concern for his wife (and perhaps by that alone).[1] Or consider Stocker's (1976) case of a friend who visits you in hospital because she believes it is her moral duty. Again, the agent appears to do what is right, but insofar as she is moved by concern for the rightness of her actions as such rather than concern for you in particular, her motivations strike us as problematic.

Whatever else we might say about moral fetishism, the following is relatively uncontroversial: the moral fetishist cares about the rightness of her actions, but in the wrong sort of way. Her concern for morality seems to exclude or overshadow any concern she has for those things that make right actions worth pursuing— things like the fact that it is one's wife one can save.

Consideration of cases like these have lead many philosophers to conclude that wanting to do what is right *as such* is invariably morally fetishistic. Sure enough, the good person wants to do what is right, they say, but their desire should be understood *de re,* not *de dicto*. The good person wants to perform those actions that are in fact right (whether or not they recognize them as such), but the content of their desire is not *to do what is right*. The good person saves their wife *because it is their wife,*

[1] As Williams notes, "[i]t might have been hoped by some (for instance, by his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife" (p. 18).

and visits their friend in hospital *because it is their friend,* not because they have a self-consciously moral desire to do what is right, whatever that might be.

This popular view—that wanting to do what is right as such is morally fetishistic—gets much of its appeal from the fact that it appears to deliver the right verdicts in cases like Williams' and Stocker's. Yet this view gets the wrong results in many cases of moral uncertainty. In cases of moral uncertainty, there is no particular action such that (1) it is right and (2) the person in question wants to do it. So if, in such a case, a person wants to do what is right (as I think it is often natural to think), their desire to do what is right must be de dicto. On the popular view, that makes them moral fetishists. But intuitively, they aren't.

Why do some cases of wanting to do what is right as such strike us as perfectly fine while others seem deeply objectionable? This is the central question of this paper. In thinking about cases like these philosophers have generally sought to vindicate one set of intuitions at the expense of the other. On the one hand there are those who claim having a de dicto desire to do what is right is invariably morally fetishistic (the view I canvassed above). On the other, there are those who claim that having a de dicto desire to do what is right is not merely unobjectionable, but is in fact characteristic of the motivations of good people, and perhaps even a necessary condition on an action's having moral worth. According to many proponents of this view, it is only in virtue of having a desire to do what is right as such that a person can be reliably disposed to perform right actions, and can said to have a genuinely good will.

In this paper, I argue for an account of moral fetishism that charts a middle way between these two views, vindicating our intuitions about the variety of cases in which people are moved by a desire to do what is right as such. What distinguishes the moral fetishist, on my view, is not that they have a de dicto desire to do what is right, but that they have an *unconditional* desire to do what is right. This means that, despite the fact that they both want to do what is right as such, the good person and the moral fetishist have desires with different contents.

If my view is right, then both the claim that it is invariably morally fetishistic to want to do what is right as such and the claim that these desires are characteristic of good moral motivation are wrong. However, I'll spend more time arguing against the former claim than the latter, largely because the former has recently been used as a crucial premise in a number of different arguments, for a wide variety of views.[2] Moreover, this claim is often cited as somewhat self-evident, receiving very little in

---

[2] To mention just a few of those arguments: the claim that wanting to do what is right is ipso facto morally objectionable is central to Weatherson's (2014) argument against 'hedging' in cases of moral uncertainty. Hedden (2016) follows Weatherson in taking the supposedly fetishistic nature of wanting to do what is right as such as the basis of an argument against the view that there is any "normatively interesting sense of ought in which what you ought to do depends on your uncertainty about (fundamental) moral facts." Markovits (2010) cites the allegedly fetishistic nature of wanting to do what is right as such as a motivation for her account of moral worth. On her view, it is a necessary condition on an action's having moral worth that the person who performs it be motivated by the reasons that make the action right, rather than the fact that the action is right. To be motivated by the fact of an action's rightness as such is, again, to have a fetishistic concern with morality as such.

the way of support save for an appeal to the kinds of cases of moral fetishism we've already seen. Though others have called both of these aforementioned views into question, they have not offered a positive account of moral fetishism that allows us to capture the motivating insights of both. That is what I attempt to do here.

The plan for the paper is as follows. In section 1 I argue that cases of moral uncertainty give us prima facie reason to be skeptical of the claim that wanting to do what is right de dicto is always morally fetishistic. In sections 2 I consider an objection to this argument which calls into question the moral character of those who are morally uncertain. In Sections 3 and 4 I consider two prominent arguments for the claim that wanting to do what is right is invariably morally fetishistic, one inspired by Smith's (1994) argument against motivational externalism, and one offered by Brian Weatherson (2014). I argue that both arguments fail. However, the reasons why they fail are illuminating, and set up my positive account of moral fetishism (section 5). I show how my account deals with the puzzle of our seemingly inconsistent intuitions about cases. I then close by considering why so few philosophers have seriously entertained the suggestion that wanting to do what is right de dicto is sometimes praiseworthy and sometimes problematic.

1. Moral Uncertainty

Pretheoretically, it seems quite natural to think that when a good person is morally uncertain, they may want to do what is right. To motivate this claim, consider the following case:

> Selina is an undergraduate who has taken an interest in the work of Prof. M., a philosopher. Selina asks Prof. M. if she would be willing to take her on as an unpaid research assistant over the summer. Though she could use the extra help, Prof. M. is not sure how she should respond. On the one hand, she wants to help Selina; Selina would undoubtedly benefit from the work experience, and she might even warrant a letter of recommendation for her upcoming graduate school applications. And it's clear to Prof. M. that Selina is a talented young philosopher. Prof. M. thinks that it is important to give encouragement, support and opportunities to underrepresented groups in the profession.
>
> On the other hand, Prof. M. worries that accepting Selina's offer of unpaid work might be exploitative. Being a research assistant is a skilled job for which one should be fairly compensated, she thinks, regardless of one's willingness to work without pay. And even if it's not exploitative, perhaps it would be unfair: unpaid research work is professionally valuable, but is not an option for anyone but the financially privileged. If every philosopher took on students happy to do this kind of unpaid work, then those who cannot afford to work for free would be at a professional disadvantage. (Prof M. thinks that this consideration alone might be a reason to not accept Selina's request, but she isn't sure.)
>
> Prof. M. wants to do what is right, and she thinks that the right thing to do is whatever is supported by the balance of these considerations. The trouble is, she just doesn't know what that is.

This is a case of genuine moral uncertainty. As it happens, it is also a case of uncertainty about non-moral facts; Prof. M doesn't know whether writing a letter of recommendation for Selina would make the difference between her being accepted to a graduate program or not, nor does she know whether declining Selina's offer would mean that Selina just takes other, less valuable unpaid work

over the summer. But even if Prof. M. did know these things, we can suppose that she still might not know what she ought to do. This additional information may not be enough for her to settle the question of whether taking on an unpaid RA would be *exploitative*, or if it contributes to an *unfair* distribution of professional resources. It also may not help her to resolve the question of how she ought to weigh those considerations against one another. To her, it simply seems that accepting Selina's offer would be good in some ways and bad in others.

Though she doesn't know what she ought to do, Prof. M. wants to do what is right. Of course, that's not all she wants; she also wants to help Selina, she wants to do what is fair, and she wants to advance the cause of women in philosophy (among other things). But because she cares about how she ought to weigh these considerations against each other, it is perfectly natural to think of her as wanting to do what is right (and indeed, it is natural for her to think of her own motivations this way, too). Ascribing to her a desire to do what is right also helps to explain why she engages in careful deliberation about what to do.[3]

As I've described the case, Prof. M.'s motivations are, I think, quite natural and morally unobjectionable. She wants to do what is right, but her motivations don't strike us strange or troubling. This presents a problem for those who claim that wanting to do what is right de dicto is invariably morally fetishistic. For as we've already seen, Prof. M.'s desire to do what is right cannot be a desire to do what is right *de re*; after all, there is no particular action such that she wants to perform it, so, necessarily, there is no particular action such that it is both right *and* she wants to perform it (whether or not she knows that it is right). Her desire to do what is right is de dicto. Intuitively, however, she is not guilty of moral fetishism. That's why this case gives us *prima facie* reason to be skeptical of the claim that it is invariably morally fetishistic to do what is right de dicto.

2. Morally Uncertain Agents are not Morally Ideal

Let us briefly recap the argument from Section 1 before turning to an objection. In cases of moral uncertainty, a person may want to do what is right. In those cases, there is no action such that 1) it is right, and 2) the person wants to do it. So their desire to do what is right must be de dicto, not de re. But wanting to do what is right in such cases often seems perfectly unobjectionable (or at least, it doesn't seem morally fetishistic). We therefore have prima facie reason to be skeptical of the claim that wanting to do what is right de dicto is invariably morally fetishistic.

One might challenge my appraisal of cases of moral uncertainty like Prof. M.'s, insisting that there really is something objectionable about the motivations of the agents concerned. After all, it is certainly true that morally *ideal* agents wouldn't find themselves not knowing what they ought to do, at least not when they have access to all of the relevant empirical facts. In thinking about cases like Prof. M's, then, we're already imagining people who fall short of some important moral ideals.

To understand the possible forms this objection might take, it will be helpful to distinguish between two features of people like Prof. M. that might occasion our concern: the first is their moral ignorance, and the second is their desire to do what is right in the face of their moral uncertainty.

[3] For an argument that *only* a desire to do what is right de dicto can explain why agents in situations of moral uncertainty engage in moral deliberation, see Aboodi (2016, 2017) and Johnson-King (2020). My argument only needs the weaker claim that there are at least *some* cases in which such a desire motivates people to engage in moral deliberation.

Neither of these features, I'll argue, plausibly make the motivations of Prof. M. or people relevantly like her morally fetishistic.

First, moral ignorance. Prof. M. doesn't know what she ought to do, and we are supposing that this is not due to mere factual ignorance. Prof. M.'s ignorance is "pure" moral ignorance. Many have argued that such pure moral ignorance impugns a person's moral character, not merely their status as knowers. Perhaps the most sustained defense of this position is from Elizabeth Harman, who argues that although we can be blamelessly ignorant on matters of fact, we are not similarly blameless when it comes to matters of morality.[4]

Suppose this is right. Then if Prof. M. does the wrong thing despite wanting to do what is right, she may be morally blameworthy for her actions. Moreover, we might even think that the mere fact that she doesn't know what she ought to do reflects poorly on her moral character. But notice that although there may be broad constraints on just how morally ignorant or mistaken one can be while nonetheless being morally good, (a person who believes it is right to promote only her own pleasure, for example, might not make the cut) it is not usually taken to be a requirement on being a generally good person that one *always* form correct moral judgements, let alone that one do so with the kind of confidence that would rule out further doubt and second-guessing. Indeed, as we saw in our discussion of Smith's argument against motivational externalism, it is widely taken to be a constraint on a moral theory that it be able to accommodate the fact that a good and strong-willed person's motivations track changes in their moral judgements over time. This means that even good people change their minds about what is right, and that they are therefore sometimes wrong.

Moreover, being morally uncertain does not mean one is completely in the dark when it comes to moral matters. Prof. M. knows, for example, that Selina's welfare is relevant to what she ought to do, and she knows that this is a fairly weighty consideration. But she thinks this consideration may be outweighed by others, like considerations of fairness. Despite her uncertainty, Prof. M. has robust, substantive conception of right action: she knows what she morally ought to do in a great variety of situations, and she has a good sense of the kinds of features of actions that matter, morally speaking. But her substantive conception of right action is insufficiently detailed to deliver a verdict on what she ought to do in every situation she might face.

In this respect, I think we are all like Prof. M. Perhaps a morally ideal agent would always know what they ought to do, but the rest of us don't. Moral reasoning can be difficult for even the most sophisticated among us. We need to figure out not only how to weigh various competing moral considerations, but also to determine, in each situation, which moral considerations apply. Is having Selina work without pay exploitative? And if it is, how should that be weighed against considerations of Selina's welfare? If moral uncertainty involves a kind of moral shortcoming, it is a shortcoming

[4] See e.g. Harman (2011). The connection between questions about the way that moral uncertainty may reflect on our moral character and questions about responsibility and blameworthiness for actions done from moral ignorance is somewhat indirect. Much of the recent work on moral ignorance and blameworthiness focuses on the question of whether moral ignorance, like non-culpable factual ignorance, can be exculpatory when someone fails to do what they morally ought to do. Is a person blameworthy for performing a wrong action if that action was the result of sincerely held, false moral views? What if that false moral view is widely held in their community? Our question does not concern a person's responsibility or blameworthiness for their actions, but whether moral ignorance itself—regardless of what actions it might lead to—reflects poorly on a person's moral character.

we all share. And yet we are not all moral fetishists. Prof. M.'s moral ignorance might make her, like all of us, morally non-ideal, but intuitively, it doesn't make her a moral fetishist.

What then of the idea that it is not moral ignorance itself, but wanting to do what is right in the face of moral uncertainty that is morally fetishistic? Prof. M.'s desire to do what is right doesn't seem to come at the expense of concern for the morally relevant features of her situation, like promoting Selina's interests, acting fairly, and avoiding the exploitation of others. In fact, it seems natural to think that it is because she cares about these things that she wants to do what is right. And Prof. M. doesn't seem to care *too much* about wanting to do what is right. Her desire to do what is right doesn't seem to saddle her with "one thought too many." And unlike Williams' protagonist, Prof. M.'s "moral thought" doesn't appear to be in any way objectionably impersonal. Given that she doesn't know what she ought to do, it seems perfectly natural for her to want to do what is right.

It's also important to keep in mind that the relevant question for our purposes is not whether Prof. M. is an ideal moral agent, but whether she has made a fetish of doing what is right. As I've described the case, I don't think that many would be tempted to think that Prof. M.'s desire to do what is right is strange, objectionable, or morally fetishistic. Given the complexity of moral reasoning and decision making, it is not surprising that good people like Prof. M. sometimes find themselves in situations of moral uncertainty, nor is it prima facie worrying that in some such situations they care about getting things right.

Given the ubiquity of moral uncertainty and the plausibility of the claim that in such cases, people may want to do what is right, why is it so often claimed that there is something wrong with wanting to do what is right as such? So far, the only reason we've seen is that this claim appears to explain what's wrong in cases like Williams' and Stocker's. But this is not the only source of its appeal. The *locus classicus* in debates on de dicto moral motivation, and the argument commonly taken to support the claim that it is morally fetishistic to want to do what is right as such, is Smith's (1994) argument against externalist theories of moral motivation.[5] How Smith's argument fares with respect to the debate between motivational internalists and externalists is not my concern here, but the phenomenon he draws our attention to is. Closer examination of that argument shows that it not only fails to give us reason to think there is always something wrong with wanting to do what is right as such, it also relies on a picture of what it is to have a de dicto desire to do what is right that does not extend to cases like Williams'.

2. Motivational Externalism and the Fetishism Objection

Smith's argument begins with an example of the kind of connection between moral judgements and motivations that he thinks an adequate moral theory must be able to explain. Smith asks us to imagine being engaged in an argument about a "fundamental moral question," namely, whether one should vote for the social democrats or the libertarian party at some election:

> [W]e will suppose that I come to the argument already judging that we should vote for the libertarians, and already motivated to do so as well. During the course of the argument, let's suppose that you convince me that I am fundamentally wrong. I

[5]Those who cite Smith's argument in support of the claim that it is morally fetishistic to want to do what is right as such include Arpaly and Schroeder (2013), Arpaly (2002), and Markovits (2010).

> should vote for the social democrats, and not just because the social democrats will better promote the values that I thought would be promoted by the libertarians, but rather because the values I thought should and would be promoted by the libertarians are themselves fundamentally mistaken. You get me to change my most fundamental values. In this sort of situation, what happens to my motives? (p.72).

The challenge is to explain why, when a "good and strong-willed" person forms new moral judgements, their motivations tend to change in step. The motivational internalist claims that the connection between moral judgements and motivations is internal to moral judgements themselves: it is in the nature of moral judgements that they motivate us to act in accordance with them. The motivational externalist denies this, claiming instead that the disposition to be moved by one's moral judgements is not only defeasible, as when one suffers weakness of will, but altogether contingent.[6]

So how can the externalist explain the reliable (though defeasible) connection between moral judgements and motivations in good people? They cannot appeal to the content of moral judgements themselves (that would amount to a form of internalism). Instead, they must appeal to some further motivational state possessed by good people. The question for the externalist, then, is what the content of this further motive must be.

Smith thinks there is only one answer available to the externalist: good and strong-willed people have a standing desire to do what is right, *whatever that might be*. This desire provides the missing piece the externalist needs; when a good person judges it right to vote for the social democrats, for example, they become motivated to do so because they have a standing desire to do whatever it is that turns out to be right.

To reiterate: the motive that the externalist must posit, according to Smith, is a desire to do what is right, where this desire is understood de dicto rather than de re. Of course, the de re and de dicto interpretations are not mutually exclusive. A good and strong-willed person who has a de dicto desire to do what is right and who also has a (we are supposing) true belief that they morally ought to vote for the social democrats, will also have a de re desire to do what is right—that is, they will desire to vote for the Social Democrats. But the problem, according to Smith, is that on the externalist's picture this de re desire must be derivative. More specifically, it must be derived from the good person's de dicto desire to do what is right in the way instrumental desires are derived from final desires.

To see this, consider again the argument between the libertarian and the social democrat. On the view that Smith takes to be unavoidable for the externalist, "when I no longer believe it is right to vote for the libertarians, I lose a derived desire to vote for them, and when I come to believe that it is right to vote for the social democrats, I acquire a derived desire to vote for them" (p.74). At no point is the agent's motivation to vote for her preferred party non-derivative. And this, Smith thinks, is a completely implausible picture of how good people are motivated.

> [I]f this is the best explanation the strong externalist can give of the reliable connection between moral judgement and motivation in the good and strong-willed person then it seems to me that we have a straightforward reductio. For the

[6] Externalists often appeal to the possibility of "amoralists"—people who make moral judgements but lack any motivation whatsoever to act in accordance with them—as evidence for their view.

> explanation is only as plausible as the claim that the good person is, at bottom, motivated to do what is right, where this read de dicto and not de re, and that is surely a quite implausible claim. For commonsense tells us that if good people judge it right to be honest, or right to care for their children and friends and fellows, or right for people to get what they deserve, then they care non-derivatively about these things. Good people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like, not just one thing: doing what they believe to be right, where this is read de dicto and not de re. Indeed, commonsense tells us that being so motivated is a fetish or moral vice, not the one and only moral virtue (1994, p. 75).

The objection here is that on the view the externalist is forced to adopt, the good person desires to help their loved ones, to keep their promises, or to do what is honest or kind, only as a *means* to doing what is right. And this, Smith rightly claims, is not what good people are like. Good people do not care only instrumentally about helping their loved ones, keeping their promises and doing what is honest or kind; they care about them for their own sake, as ends in themselves.

As Smith notes, his concern here is strikingly similar to Williams' cited earlier: both arguments trade on a kind of skepticism towards those who are moved by "self-consciously moral motives." These motives furnish the agent with "one thought too many," and they "alienate" her "from the ends at which morality proper aims" (p. 76). But where the idea of alienation is left somewhat obscure in Williams' argument, Smith spells it out in terms of derivative (i.e. instrumental) desire: a person is alienated from the proper ends of morality insofar as she cares about them only as a means to doing what is right as such.

With Smith's argument now on the table, the first thing to notice is that although it is often cited in support of the claim that it is morally fetishistic to want to do what is right de dicto, Smith's argument doesn't establish or even rely on this claim. His argument relies instead on a much more specific (and intuitively plausible) claim: that there is something wrong— something morally fetishistic— if one's concern for honesty, the well-being of one's fellows and the like is merely instrumental to one's final end of doing what is right. This more specific claim does not entail the more general one; one might grant that there really is something wrong with caring about honesty and the like only instrumentally, but deny that having a de dicto desire to do what is right thereby means that one cares about such things only instrumentally. And Smith doesn't give us an independent reason to think that if one has a de dicto desire to do what is right, then one's concern for things like honesty and the well-being of their fellows must be instrumentally derived from that desire.

What Smith's argument gives us is, at best, a characterization of what a person who wants to do what is right must be like *if their desire is to play the role (Smith thinks) the externalist needs it to play*. But this leaves it open for the motivational internalist, or anyone else, to claim a more modest role for the desire to do what is right among the motivations of good people.[7]

---

[7] It should also be noted that, as it applies to the motivational externalist, Smith's argument is only as strong as his claim that the aforementioned strategy of appealing to underlying "self-consciously moral motives" is

The second thing to notice about Smith's argument is that it doesn't provide us with a general account of moral fetishism. Even if we agree with Smith that having a merely instrumental concern for things like honesty, the well-being of one's fellows, people getting what they deserve, etc. is a kind of moral fetishism, that can't be all there is to it. After all, the diagnosis of moral fetishism as a kind of merely instrumental concern doesn't comfortably fit what's going on in William's case—a paradigm example of moral fetishism.[8] Granted, it is not clear how exactly to interpret the motives of Williams' protagonist. According to Williams', the man has 'one thought too many.' But is the worry here that without that further, moral thought, the mere fact that it was his wife would not have been sufficient to motivate him? Or is the concern that even if this fact were sufficient to motivate him, there is still something wrong with his caring about the moral permissibility of his action? Williams' doesn't say. But however we interpret the motivating role of this extra, moral thought, we can reasonably suppose that Williams' man *does* care about his wife and he does want to save her, and not merely because he believes that this is a means or a way of doing what's right. Nonetheless, his motivations still strike us as strange and troubling. He may care nonderivatively about saving his wife, but he seems to care about the moral permissibility of his actions even more. Why, we might wonder, is his desire to save his wife potentially overrideable? And why does he care so much about doing what is morally permissible if he thinks that what is permissible might conflict with saving his wife? The problem, it seems, is not so much that he cares about his wife in the wrong sort of way, but that he cares about doing what is right in the wrong sort of way.

Where does all of this leave us? To be motivated in the way Smith describes does seem like a form of moral fetishism. But Smith's argument doesn't give us a reason to think that there is, in general, something wrong with wanting to do what is right as such. And as we saw in the case of Prof. M., wanting to do what is right as such can sometimes seem perfectly unproblematic. Moreover, we don't yet have a good, general picture of what's wrong with wanting to do what is right in those cases like Williams' when it really *does* seem wrong.

And so we have the puzzle that motivates this paper: why are some cases of wanting to do what is right as such perfectly unproblematic, while other are deeply objectionable? And what is moral fetishism?

3. Moral fetishism and the Argument Against Hedging

Like Smith's argument against motivational externalism, Weatherson's (2014) argument against "moral hedging" appeals to the idea that it is morally fetishistic to be merely instrumentally motivated by those features of actions that make them right. Unlike Smith, however, Weatherson explicitly argues that *all* cases of wanting to do what is right de dicto involve being motivated in this

---

the externalist's best bet for explaining the motivational dispositions of good people. But as Dreier (2000) has argued, there may be other, externalist-friendly ways of doing this same work that avoid altogether the need to posit such motives.

[8] In all fairness, Smith does not claim to give an account of moral fetishism, or to diagnose what is driving our intuitions in Williams' case. He merely claims that "The present objection to externalism is like Williams' objection… the objection in this case is simply that, in taking it that a good person is motivated to do what she believes right, where this is read *de dicto* and not *de re*, externalists too provide the morally good person with "one thought too many." They alienate her from the ends at which morality proper aims" (p. 75-6). The success of his argument against the externalist depends on his having identified just one form of moral fetishism, and this is consistent with Williams' example requiring a very different treatment.

way. It is therefore worth looking turning to his argument to see if it supports the general claim that it is invariably morally fetishistic to want to do what is right as such. It will help to begin with one of Weatherson's central examples of moral hedging:

> Martha is deciding whether to have steak or tofu for dinner. She prefers steak, but knows there are ethical questions around meat-eating. She has studied the relevant biological and philosophical literature, and concluded that it is not wrong to eat steak. But she is not completely certain of this; as with any other philosophical conclusion, she has doubts. As a matter of fact, Martha is right in the sense that a fully informed person in her position would know that meat-eating was permissible, but Martha can't be certain of this. What should she do? (p. 143).

According to Weatherson, cases of moral hedging have the following structure. An agent is deciding whether to Φ (eat the steak), but is uncertain whether Φing is morally permissible. The agent is uncertain about the permissibility of Φing because she is uncertain about which of two moral theories is true. The agent believes that theory A is probably true, and theory A says that Φing is morally permissible. But she thinks there is a small chance that theory B is true instead, and theory B says that Φing is seriously wrong. The agent hedges, morally, just in case she decides not to Φ because Φing *might* be seriously wrong (even though it probably isn't).

Weatherson's claim is that Martha need not hedge her bets and stick to the tofu. Given that a fully informed person in Martha's position would know that meat-eating is permissible, it would not be wrong for Martha to be "morally reckless" and eat the steak. His argument here is part of a broader project defending the view that "the most important norms concerning the guidance and evaluation of action and belief are external to the agent being guided or evaluated" (p. 141).

Whether Martha ought to take the safe option and hedge, or whether there is something wrong with moral hedging in general, is not my interest here. My interest is rather in Weatherson's stated reason for thinking there is something wrong with moral hedging.

The problem with moral hedging, Weatherson thinks, is that to hedge, one must be motivated by a concern with "morality de dicto. And that seems wrong" (p. 152). But does it seem wrong? It's certainly not clear that if Martha were to hedge and choose the tofu, her motives would strike us as problematic or somehow objectionable. But perhaps they *should* strike us as problematic. In discussing the case of Martha, above, he says "Why should she turn down the steak? Not because she values the interests of the cow over her dining. She does not" (p. 152). Weatherson doesn't tell us what Martha thinks makes actions right, but he does tell us that she doesn't think that the cow's interests have anything to do with it (though, again, she can't be certain). So if she is moved to eat the tofu by a desire to do what is right, then she isn't moved by a non-instrumental concern for the interests of the cow, nor is she moved by whatever she does think makes actions right; after all, by her lights, eating the steak is morally fine. Instead, she must be moved by a concern for rightness, *whatever that might be.*

Weatherson's claim here is that when agents hedge between competing moral theories, they are not moved by non-instrumental concern for those features of actions that they think make actions right. They are moved instead by the thought that these features might matter morally. But to care about morality independently of those features that make right actions worth performing is a kind of moral fetishism.

This objection to moral hedging relies on a claim about the kind of desire that must be involved in moral hedging: it is a desire to do what is right de dicto that is independent of any non-instrumental concern for the lower-order, right-making features of actions. I suspect that Weatherson thinks that this kind of concern for rightness de dicto—the kind that is independent of a concern for the lower-order, right-making features of actions—is just what wanting to do what is right de dicto amounts to, both in contexts of moral hedging and outside of such contexts. And this is why he thinks that wanting to do what is right de dicto is invariably morally fetishistic.

As I hope will by now be clear, I think this is mistaken. As I've presented the case, Prof. M. is not engaged in moral hedging (my description of the case leaves open how she chooses to act, and on what grounds), but she does want to do what is right de dicto. And her desire to do what is right is not independent of her concern for those things that she thinks make actions right. In fact, the most natural explanation for why she wants to do what is right is that she has a non-instrumental concern for the various moral considerations that she thinks are at stake. She wants to get things right, morally speaking, because she wants to help those from historically underrepresented groups in philosophy, she wants to avoid exploiting others, she wants to do what is fair, and she believes that the right action is the one that is favoured by the right balance of these considerations.[9]

What about Martha? Is she guilty of moral fetishism in virtue of moral hedging? Martha is quite sure that meat-eating is morally permissible, and she assigns only a low probability to the theory which says that the interests of the cow are morally significant. This is why Weatherson claims that she does not have a non-instrumental concern for the interests of the cow. But this need not mean that her desire to do what is right is utterly independent of those things that, according to her own moral view, make actions right. After all, in worrying about the risk that she is doing something seriously morally wrong, she may simply be worried that eating the steak is wrong by the lights of her own moral theory, if it were fully worked out. That is, she might be worried that, given her *current* commitments and values (which presumably include things like a concern for the suffering of other people, and creatures relevantly like them) she *should* care about the interests of the cow. If this is right, then it is not at all clear that her desire to do what is right is morally fetishistic.

The upshot is that, whether it motivates an agent to morally hedge or not, a desire to do what is right as such need not be independent of those things that the agent thinks make actions right. People like Prof. M. (and maybe even Martha) may want to do what is right precisely because they care about the various morally relevant things at stake.

This insight brings into relief a common thread running through both Smith and Williams' discussions of moral fetishism. Though they provide different diagnoses of the problem—Smith identifying moral fetishism with a kind of merely instrumental concern, and Williams' with a kind of alienation from one's ground projects—they both identify ways that a person's desire to do what is right can fail to be sufficiently explained by their other values and commitments. One need not deny Smith's insight that

[9] Weatherson anticipates an objection that the "strong form of Smith's fetishism objection" on which he relies can't explain why moral or prudential reflection is a good practice. In response, he says "as long as we accept that there are genuinely plural values, both in moral and prudential reasoning, we shouldn't think that a desire to do what is right is driven by a motivation to do the right thing, or to live a good life, as such" (p.161). This suggests that Weatherson would not consider someone like Prof. M. as having a desire to do what is right as such. If that is right, then what he means by "a desire to do what is right as such" must be something other than a desire to do what is right de dicto (which Prof. M. undoubtedly has).

being merely instrumentally concerned about fairness and kindness and the like is a form of moral fetishism, nor deny that there is something wrong with one's motivation to save one's wife being contingent on its moral permissibility. But neither of these cases give us reason to believe that there is something wrong with wanting to do what is right per se. Rather, as I will go on to argue, they give us reason to think that in assessing an agent's motivations, we need to pay attention to the structure of the agent's desires—how their various motivations support and explain each another.

4. What is Moral Fetishism?

What, then, is it that makes someone a moral fetishist? On my view, to be a moral fetishist is to want to do what is right de dicto, whatever rightness might be. We can make this thought more precise by appealing to the notion of a conditional desire.

A conditional desire has the form: I want p provided that condition C is satisfied. Perhaps the most familiar kind of conditional desires are those that are conditional on their own persistence.[10] For example, my desire for a beer later is not satisfied if, later, I get a beer but no longer want it.[11] But a desire can be conditional on things other than its own persistence. I want to go to the art gallery tonight, but only on the condition that they are showing the Dutch Masters. If they aren't showing the Dutch Masters, my desire to go the gallery won't be satisfied, even if I go. The conditions in each case qualify *what* the agent wants. I want a beer later, but not if, later, I no longer want a beer. I want to go the gallery, but not if they're only showing Jeff Koons.

What does it mean to say that the moral fetishist wants to what is right, whatever rightness might be? It means that the moral fetishist has an unconditional desire to do what is right de dicto.

When a good person wants to do what is right de dicto, her desire to do what is right is implicitly conditional on what morality requires of her and (more importantly) why. Prof. M. wants to do what is right, but only on the condition that the right action is one that is favoured by the balance of the considerations that she believes are morally important. These include helping underrepresented groups in philosophy, respecting others' labor, acting fairly, and the like. The moral fetishist's desire to do what is right, by contrast, is unconditional. Her desire to do what is right is independent from her substantive conception of right action.

More precisely, the good person's desire to do what is right de dicto (on a particular occasion) is conditional on her having reasons to do what is right that count as reasons from the point of view of her own substantive conception of morality. Consider a version of Kant's famous example of the murderer at the door. One day, Cain comes to your door, and asks if you know where Abel is. You happen to know that Cain intends to kill Abel. You're not entirely sure what you ought to do. Should you lie to Cain? Persuade him to turn himself in? Something else? You want to do what is right, but you're not sure what that is. So, your desire to do what is right must be de dicto. But your desire to do what is right is also conditional. You do not want do what is right *even if* acting rightly requires disclosing Abel's location. Why not? From your point of view, even if we are morally obligated to tell the truth to others, that obligation couldn't be strong enough to require doing something that gets Abel killed. If getting Abel killed to avoid lying is right, you'd rather do what's wrong.

[10] Here I am following Derek Parfit (1984) p. 151.
[11] This example is also offered by McDaniel and Bradley (2008), p. 267

What about the moral fetishist? Like the good person (you), the moral fetishist wants to do what is right, but isn't sure what that is. The difference is that her desire to do what is right is not similarly conditional. If telling the truth turns out to matter more, morally, than preventing Abel's death, then so be it. Unlike the good person, the moral fetishist is indifferent to what morality is like. She wants to do the right thing, *no matter what acting rightly requires, or why*. And that is what makes her a fetishist.

So, to recap: to be a moral fetishist just is to have an unconditional desire to do what is right de dicto. Wanting to do what is right de dicto is consistent with being a good person, provided your desire is conditional on your having reasons to do what is right that count as reasons given your substantive conception of right action.

One thing this shows is that, while it appears that the moral fetishist and the good person have the very same desire, in fact they want different things. The good person's desire to do what is right de dicto is conditional, the moral fetishist's unconditional.

This is an example of desire underspecification. As Delia Graff Fara and others have observed, a desire report can be true without fully specifying what it is that the agent wants.[12] Consider an example from Fara (2013):

> Fiona says that she wants to catch a fish; Charlotte says that she wants to have some champagne. Neither has expressed with full specificity what it is that she wants. Fiona wants to catch a fish that's big enough to make a meal; a minnow will not do. Charlotte wants enough champagne to feel it go to her head; a thimbleful will not do. Nevertheless, each speaks truly (p. 250).

As the foregoing discussion shows, knowing that an agent wants to do what is right de dicto does not fully specify what it is that they want. To know whether they are a moral fetishist, it is not enough to know that they have this desire. One needs to know, further, what conditions (if any) apply to their desire. And this requires knowing how their desire to do what is right is connected, motivationally, to their conception of what matters.

The fact that desire reports underspecify the satisfaction conditions of desires helps to explain an otherwise curious feature of debates about whether wanting to do what is right de dicto is morally fetishistic. The literature is dominated by two diametrically opposed views. As we have seen, many philosophers argue that wanting to do what is right de dicto is always morally objectionable. But others (often working in the Kantian tradition) say the opposite: being good *requires* being motivated to do what is right de dicto.[13] The middle road I have taken here, on which wanting to do what is right de dicto is sometimes laudable and sometimes objectionable, has had few takers. This is surprising, given the view's considerable intuitive appeal.

We are now in a position to see why the appealingly irenic middle way I propose here has been largely ignored. It is natural to assume that any two people who want to do what is right de dicto thereby want the same thing. And if this desire provides their motivating reason for action, then their motivations must be morally on a par. Once we attend to the phenomenon of desire

[12] See Fara (2003, 2013), Lycan (2012), and Grant and Phillips-Brown (2019).
[13] The claim is often put in terms of a necessary condition on an action's having moral worth. Those who hold this view include Sliwa (2015a, 2015b), and Johnson-King (2018).

underspecification, however, it becomes clear that this need not be the case. We can draw finer distinctions among those who want to do what is right de dicto, and doing so allows us to see that only some of these individuals fetishize morality. A person only fetishizes morality if their concern for doing the right thing is disconnected from their concern for what they think really matters.

References

Aboodi, R.(2016). "The Wrong Time to Aim at What's Right: When is De Dicto Moral Motivation Less Virtuous?." *Proceedings of the Aristotelian Society* 115(3), 307-314.
Aboodi, R. (2017). "One Thought Too Few: Where *De Dicto* Moral Motivation is Necessary." *Ethical Theory and Moral Practice* 20(2), 223-237.

Arpaly, N. (2002). *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford: Oxford University Press.

Arpaly, N. and Schroeder, T. (2013). *In Praise of Desire*. Oxford: Oxford University Press.

Dreier, J. (2000). "Dispositions and Fetishes: Externalist Models of Moral Motivation." *Philosophy and Phenomenological Research* 60(3), 619-638.

Fara, D. G. (2003). "Desires, scope, and tense." *Philosophical Perspectives* 17(1), 141–163. (Originally published under the name "Delia Graff".)

Fara, D. G. (2013). "Specifying Desires." *Nous* 47(2), 250-272.

Fried, C. (1970). *An Anatomy of Values*. Cambridge, MA: Harvard University Press.

Grant, L. and Phillips-Brown, M. (2019). "Getting What You Want." *Philosophical Studies* (currently available online only).

Harman, E. (2011). "Does Moral Ignorance Exculpate?" *Ratio* 24, 443-468.

Hedden, B. (2016). "Does MITE Make Right? Decision-Making Under Normative Uncertainty." In R. Shafer-Landau, ed., *Oxford Studies in Metaethics*, vol. 11. Oxford: Oxford University Press.

Johnson King, Z. A. (2018). "Accidentally Doing the Right Thing." *Philosophy and Phenomenological Research*.

Lycan, W. (2012). "Desire considered as a propositional attitude." *Philosophical Perspectives* 26(1), 201–215.

Markovits, J. (2010). "Acting for the Right Reasons." *Philosophical Review* 119(2), 201-242.

McDaniel, K. and Bradley, B. (2008). "Desires." *Mind* 117, 267-302.

Parfit, D. (1984). *Reasons and Persons*. New York: Oxford University Press.

Sliwa, P. (2015a). "Praise without Perfection: A Dilemma for Right-Making Reasons." *American Philosophical Quarterly* 52(2).

Sliwa, P. (2015b). "Moral Worth and Moral Knowledge." *Philosophy and Phenomenological Research* 93(2), 393-418.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

Stocker, M. (1976). "The Schizophrenia of Modern Ethical Theories." *The Journal of Philosophy* 73(14), 453-466.

Weatherson, B. (2014). "Running Risks Morally." *Philosophical Studies* 167(1), 141-163.

Williams, B. (1981). *Moral Luck*. Cambridge: Cambridge University Press.

Chapter 2

Getting What You Want

Co-authored with Milo Phillips-Brown[*]

1. Introduction

A widely shared sentiment, articulated by Dennis Stampe, is that desire satisfaction is "truth by a different name" (1986, p. 154). The sentiment can be sharpened by appeal to two principles, one about belief and the other about desire:

> *Truth-is-Truth Principle*
> If A believes p, then A has a belief that is true in exactly the worlds where p is true.[1]
>
> *Satisfaction-is-Truth Principle*
> If A wants p, then A has a desire that is satisfied in exactly the worlds where p is true.[2]

The Truth-is-Truth Principle is true. But, we will argue, the Satisfaction-is-Truth Principle is not. An agent may want p without having a desire that is satisfied in exactly the worlds where p is true—in particular, without having a desire that is satisfied in every world where p is true. Such an agent has a desire whose satisfaction conditions are what we call *ways-specific*: it is satisfied only when p obtains in certain ways.

(The Satisfaction-is-Truth Principle presupposes that desire is a propositional attitude.[3] Whether this presupposition is true is orthogonal to our argument, which works just as well against a version of the Satisfaction-is-Truth Principle that doesn't mention propositions: if A wants to φ, then A has a desire that is satisfied in exactly the worlds where she φs.)

---

[*] The authors contributed equally.

[1] Though widely accepted, Bach (1997) questions a principle in this vein.

[2] See e.g. Searle (1983), ch. 2, Whyte (1991), Stampe (1994), Heathwood (2006). Condoravdi and Lauer (2016) give a contextualist take on the principle. Braun (2015) endorses a similar principle, which he calls "The Weak Content-Specification Version of the Relational Analysis of Desire Ascriptions" (on which more in §10): "If N is a proper name and S is an infinitival phrase (with or without explicit subject), then: if ⌜ N wants S ⌝ is true, then the referent of N has a desire that is satisfied in exactly those worlds in which the proposition that S semantically expresses is true" (p. 149).

[3] A presupposition contested by e.g. Montague (2007) and Moltmann (2013).

Consider a case. Millie says that she wants to drink milk. Suppose (and we'll revisit this supposition later) that she is right. Intuitively, Millie nonetheless does not have a desire that is satisfied when she drinks spoiled milk. Millie wants to drink milk, but, intuitively, not just any old milk will do.

To show that a case like Millie's is a counterexample to the Satisfaction-is-Truth Principle, we need to establish two claims. First, agents like Millie do want what they say they want—e.g. Millie does want to drink milk. Second, Millie indeed does not have a desire that is satisfied when she drinks spoiled milk, and similarly for agents like her.

Fara (2003, 2013) and Lycan (2012, ms) accept similar claims on the basis of similar cases.[4] We provide new arguments for both claims. Our arguments for the first go beyond those offered by Fara and Lycan for analogues of our first claim.

The only support they offer for claims analogous to our second claim is intuitions about when agents get what they want—e.g. the intuition that Millie doesn't have a desire that is satisfied when she drinks spoiled milk.[5] As you might expect, these intuitions have been contested (by Braun (2015) and Prinz (ms), as cited in Lycan (2012, pp. 205-6)). These contested intuitions about getting what you want play no role in our argument. Instead, we argue by appeal to the *dispositional role* of desire. Because agents are disposed to satisfy their desires, an agent's dispositions provide important evidence about the satisfaction conditions of her desires. That evidence, we argue, shows that desire satisfaction is indeed ways-specific.

2. The argument

Here is our argument at a high level: agents are disposed to satisfy their desires; desire-based dispositions are ways-specific; so, desire satisfaction is ways-specific.

To begin, let's fill out the case of Millie and the spoiled milk. Millie is eating a chocolate chip cookie, and says out loud to no one in particular, "I want to drink some milk, but the milk in the refrigerator is spoiled." Although her path to the refrigerator is clear, Millie does not drink the spoiled milk. We'd like to suppose that Millie really does want to drink milk, and that she is not disposed to drink the spoiled milk. In §3–5, we'll discuss whether these are legitimate suppositions—whether the case as we suppose it to be really is possible. For now, we'll assume that the suppositions are legitimate: Millie wants to drink milk and she is not disposed to drink the spoiled milk.

---

[4] Fara (2013) rejects a principle closely related to the Satisfaction-is-Truth Principle, which she calls the "content-specification version of the relational analysis" (p. 254) of desire ascriptions. She gives only an instance of the principle: "'Lora wants to be in London' is true just in case Lora has a desire that is satisfied in exactly those possible worlds in which she is in London" (p. 254) (in her (2003), she rejects a similar principle). The left-to-right direction of the principle—the direction that she objects to—is an instance of the Satisfaction-is-Truth Principle if we accept, as we should, that if Lora wants to be in London, then "Lora wants to be in London" is true. See more in §10. Lycan isn't explicit about just what principles he objects to. We read him (2012, pp. 206–7; ms, pp. 2–3) as committed to the possibility of cases that would falsify the Satisfaction-is-Truth Principle. And in his (ms), he cites Fara's (2013) and seems to side with her (pp. 2–3).

[5] Van Rooij (1999) and Persson (2005, ch. 10) also discuss these intuitions.

Millie wants to drink milk, but she isn't disposed to drink the spoiled milk—she isn't disposed to drink the only milk that she believes is available to her. It's not that she isn't disposed to drink any kind of milk at all. She is. It's rather that her disposition to drink milk is discriminating. It is *specific* to certain kinds of milk. Not just any old milk will do.

Millie has what we call a *ways-specific desire-based disposition*. If an agent has a ways-specific desire-based disposition, then for some p, (i) she wants p; (ii) there are ways for p to obtain that she is disposed to bring about; but (iii) there are other ways for p to obtain that she is not disposed to bring about, even if she believes that she can only bring it about that p obtains in those ways. Because Millie's disposition is specific to certain ways of its being the case that she drinks milk—ways in which she drinks certain kinds of milk—it is ways-specific in just this sense.

To run our argument, we need to state carefully the thesis that agents are disposed to satisfy their desires. Here's how others have stated the thesis:

> [T]he primitive sign of having a desire is trying to satisfy it. (Humberstone (1990, p. 107), riffing on Anscombe)
>
> [T]he actions a desire is a disposition to perform are those that would satisfy that desire provided the agent's operative beliefs were true. (Stampe, 1994, p. 246)
>
> [A] desire is manifested in…behaviour aimed at satisfying the desire. (Hyman, 2014, p. 85)

In stating the thesis ourselves, we commit only minimally on further questions concerning how desires relate to dispositions. We do not assume, for example, that desires *are* dispositions. And, as far as we're concerned, the principle can be contingent, or restricted to certain kinds of agents.[6] We propose:

> *Satisfaction–Disposition Principle*
> If A has a desire that is satisfied in exactly the worlds where p is true, then A is disposed to do what she believes will bring it about that p obtains.[7]

Now the argument.

P1. If Millie has a desire that is satisfied in exactly the worlds where she drinks milk, then Millie is disposed to do what she believes will bring it about that she drinks milk.(Instance of the Satisfaction–Disposition Principle)

P2. Millie wants to drink milk.

---

[6] It needn't apply, for example, to agents incapable of action, like Strawson (1994, ch. 10)'s "Weather watchers."

[7] A weaker version of this principle that employs an "other things equal" clause to accommodate troublesome cases would work just as well for our purposes, as we explain in §6.

P3. Millie is not disposed to do what she believes will bring it about that she drinks milk—she is not disposed to drink the spoiled milk.

C1. Millie does not have a desire that is satisfied in exactly the worlds where she drinks milk. (By P1 and P3)

C2. Millie wants to drink milk and Millie does not have a desire that is satisfied in exactly the worlds where she drinks milk. (By P2 and C1)

C2 is a counterexample to the Satisfaction-is-Truth Principle, which entails that if Millie wants to drink milk, then she has a desire that is satisfied in exactly the worlds where she drinks milk.[8]

In its basic form, our argument then is this: agents are disposed to satisfy their desires (P1); desire-based dispositions are ways-specific (P2 and P3); so, desire satisfaction is ways-specific (C2).

Now we'll defend the premises.

3. In defense of P2: on saying something false but helpful

In defending the premises, we claim first that *a certain principle is true*—the Satisfaction–Disposition Principle (P1). We claim second that *a certain kind of case is possible*—one where Millie wants to drink milk (P2) and isn't disposed to drink the spoiled milk, despite believing it's the only milk available to her (P3).

In arguing for P2 and P3, then, we are arguing for the possibility that P2 and P3 are true together. In this section and the next, we are concerned with defending P2. We'll assume that P3 is true and maintain that it's possible for P2 to be true as well. In §5, we'll assume that P2 is true and maintain that it's possible for P3 to be true as well.

Turn now to the argument for P2. Millie, recall, asserts that she wants to drink milk. Suppose that Millie speaks sincerely and is as good as anyone at knowing what she wants. The default position here should be that Millie does want to drink milk. That is, after all, how things would seem if you were faced with someone like Millie, who gives a sincere, well-informed report of what she wants.

---

[8] The Satisfaction-is-Truth Principle says that if A wants p, then A has a desire that is satisfied in exactly the worlds where p is true. So, strictly speaking, C2 is a counterexample to the Satisfaction-is-Truth Principle just in case the proposition denoted by the complement of "want" in "Millie wants to drink milk" is one that's true in exactly the worlds where Millie drinks milk (for more see §10). Of course it seems to be such a proposition that's denoted! (It is not, for example, the proposition that Millie drinks milk or stubs her toe.) You might worry, though, that in fact it's a different proposition. We defer here to Fara (2013), who argues extensively that the complements of desire ascriptions like "Millie wants to drink milk" do denote the propositions that they seem to.

(To be totally clear: in maintaining that it's true that Millie wants to drink milk, we don't mean to implicate that it isn't also true that Millie wants to drink *fresh* milk. Indeed, we think it's both true that Millie wants to drink milk and true that Millie wants to drink fresh milk!)

An imaginary interlocutor might resist our claim that it's possible that Millie wants to drink milk (while not being disposed to drink the spoiled milk). The interlocutor would then need a hypothesis about why it's so natural to think that Millie does want to drink milk. Below is one such hypothesis; in the next section we consider another.

Often we say things that are false because a falsehood is most helpful for what we're trying to communicate (see e.g. Lasersohn (1999)). Take a case adapted from Sperber and Wilson (1985). Brigitte lives in Issy-les-Moulineaux, which is just outside the city limits of Paris. At a party in London, Brigitte is asked where she lives. She replies:

(1) [Brigitte:] I live in Paris.

(1) is false, since Brigitte lives just outside the city limits of Paris. Nonetheless, (1) serves its communicative purpose perfectly well.

The hypothesis is that when Millie asserts (2) she is just like Brigitte: she says something false but helpful.

(2) [Millie:] I want to drink milk.

Millie is *un*like Brigitte though. Here's why.

Brigitte must *retract* (1) in the face of the truth. Suppose that you hear Brigitte and say:

(3) [You:] Actually, Brigitte doesn't live in Paris. (She in fact lives in Issy-les-Moulineaux, which is outside of Paris.)

If Brigitte is pressed—which is it, in Paris, or just outside the city limits?—she'd be under pressure to retract:

(4) [Brigitte:] You are right; I don't live in Paris.

Brigitte must retract her original statement because one can't both live in Paris and outside of Paris (assuming one lives in just one place).[9]

But Millie does not need to retract (2) under pressure. Suppose that you hear Millie and say:

(5) [You:] Actually, Millie doesn't want to drink milk. (She in fact wants to drink fresh milk.)

If Millie is pressed—which is it, milk, or fresh milk?—she isn't under pressure to retract. She does *not* have to say:

(6) [Millie:] You are right; I don't want to drink milk.

While it can't both be true that one lives in Paris and true that one lives outside of Paris, it *can* both be true that one wants to drink milk and true that one wants to drink fresh milk. And, again, that is

[9] Yablo (2014, ch. 5) makes a similar point.

exactly what we say about Millie: it's true that she wants to drink milk, and it's true that she wants to drink fresh milk.

We can further bring out the dissimilarity between Millie's and Brigitte's cases by considering a third case, one in which the speaker says nothing false. Suppose that Yannick lives in the Marais, which *is* in Paris. At a party in London, Yannick is asked where he lives.

(7) [Yannick:] I live in Paris.

Suppose that you hear Yannick and say:

(8) [You:] Actually, Yannick doesn't live in Paris. (He in fact lives in the Marais, which is in Paris.)

This is nonsense! Yannick is under no pressure at all to retract (8). It's true that he lives in the Marais *and* it's true that he lives in Paris. Yes, Yannick could give you more information about where he lives by saying (9) instead of (7):

(9) [Yannick:] I live in the Marais.

But just because the one statement is more informative than the other does not make the first false.

The same goes for Millie. Yes, she could give you more information about what she wants by saying (10) instead of (2):

(10) [Millie:] I want to drink fresh milk.

But, again, just because the one statement is more informative than the other doesn't make the first false.

To summarize. Brigitte says one false but helpful thing (she lives in Paris) and one true thing (she lives just outside of Paris). Yannick says *two* true things, one of them (he lives in Paris) less informative than the other (he lives in the Marais). We say that Millie is more like Yannick than like Brigitte: Millie says two true things, one of them (she wants to drink milk) less informative than the other (she wants to drink fresh milk).

The analogy between Yannick and Millie is imperfect. While living in the Marais entails living in Paris, it's controversial whether wanting to drink fresh milk entails wanting to drink milk.[10]

However, our point remains: saying that Millie wants to drink milk doesn't specify everything about what she wants, just as saying that Yannick lives in Paris doesn't specify everything about where he lives. It's nonetheless true that Yannick lives in Paris. Likewise, we claim, it's nonetheless true that Millie wants to drink milk. A desire report need not be maximally specific in order to be true. Millie doesn't fully specify what she wants, but nevertheless what she says is true.

The dialectic in this section has been this. Supposing that Millie is not disposed to drink the spoiled milk, we've argued that it's possible that P2 is true—that Millie wants to drink milk. Our imaginary

[10] Heim (1992), for example, says that it doesn't, while von Fintel (1999) says that it does (see more in footnote 27).

interlocutor contested this, hypothesizing that it must be that Millie said something false but helpful. As we've seen, though, this hypothesis fails.[11]

Millie's case could of course be filled out so that she does not want to drink milk. But it clearly makes sense, and in fact seems most natural, to take Millie at her word.

4. In defense of P2: on saying and asserting

In this section we consider a different hypothesis about why it's so natural to think that Millie wants to drink milk even if, as our imaginary interlocutor argues, Millie doesn't in fact want to. This hypothesis co-opts a distinction made by Braun (2015) between what one *says* and what one *asserts*.

According to Braun, you can say a certain proposition while at the very same time asserting various other propositions. Suppose you say p and p is false. When you say p, you may at the very same time be asserting some other proposition that is true. In such a case you said something false while asserting something true. In Braun's terminology, you have *spoken truly while saying something false* (see e.g. his p. 157).[12]

If Braun is right, then the following case is possible. Millie does not want to drink milk but says that she does. When saying that she wants to drink milk, she asserts some other proposition that is true—say, the true proposition that she wants to drink fresh milk. Our imaginary interlocutor could hypothesize that *this* is why it's so natural to think that Millie says something true when she says that she wants to drink milk, even if she does not in fact want to.

There are two ways resist this thought. The first would be to deny Braun's distinction between saying and asserting. Some may deny this, but we won't try to adjudicate the issue here.

The second way is to grant Braun's distinction, but resist our imaginary interlocutor's hypothesis. This is what we'll do, maintaining that Millie's case as we've described it is unlike the kind of case that Braun cites as a "plausible example" (p. 157) of an agent using a desire ascription to assert something true while saying something false.[13]

Braun gives the following example (p. 157):

---

[11] As we noted in the introduction, Fara (2003, 2013) and Lycan (2012, ms) also argue that seemingly true desire ascriptions, like (2), are indeed true.

[12] As precedents for his view, Braun cites similar distinctions made by Bach (1994, 2001, 2005) on saying and implic-*i*-ing; Soames (2005, 2008) on semantic content and asserting; and Braun (2011) on locuting and asserting.

[13] We should emphasize that Braun is *not* committed to saying that Millie's case, as we've described it here in §4, is like his plausible example. More generally, we are not objecting to Braun's views about language: we neither object to his saying–asserting distinction (as we noted), nor do we object to the argument in which he puts that distinction to use. Rather, what we object to is the argument of an imaginary interlocutor who co-opts Braun's distinction. (See more in footnote 15 on the relationship between Braun's argument and our own.)

(11) [Suppose that Sara is teaching a philosophy seminar and suppose she has noticed that many of her students in her seminar arrived late. So she utters:] I want everyone to arrive on time for the next meeting of this seminar.

Braun invites us to suppose, following Bach (2000) and Soames (2005, 2008), that "everyone" is never contextually restricted, that it always quantifies over all people in the universe. According to Braun, what Sarah *says* is the proposition that she wants every human in the universe to arrive on time for the next seminar meeting, but she *asserts* all at once various other propositions—among them the true proposition "that Sarah wants everyone *to whom she is speaking* to arrive on time for the next meeting" (p. 158; emphasis in the original).[14] What she says is false (she does not want every human in the universe to arrive on time for the next meeting), but she nevertheless asserts a true proposition.

On our interlocutor's hypothesis, Millie is like Sarah. When Millie's dispositions are as we have supposed and she says that she wants to drink milk, she says something false but nonetheless asserts a true proposition, the proposition (say) that she wants to drink fresh milk.

But Millie is unlike Sarah, and retraction data again provide key evidence. Consider that if you insisted that Sarah doesn't really want *everyone* to come, she would be under pressure to retract, to disavow the proposition that she said. Take the following exchange, for example:

(12) [You:] Sarah doesn't want *everyone* to come to the next meeting on time! She just wants those to whom she was speaking to come to the next meeting on time!

(13) [Sarah:] Okay, fine. I don't want *everyone* to come; I just want those to whom I was speaking to come.

But as we saw in the last section, if you insisted that Millie doesn't really want to drink *milk*, she wouldn't be under pressure to retract.[15]

To summarize: we've claimed that it's possible that P2 is true—that Millie wants to drink milk, while assuming that she is not disposed to drink the spoiled milk. Our imaginary interlocutor contested this possibility, claiming that Millie said something false while nonetheless asserting something true. And while we may be able to imagine a version of our case in which this is in fact so, our interlocutor is committed to saying that if Millie is not disposed to drink the spoiled milk, she *must* be saying something false. This is what we deny.

5. In defense of P3: against the other desires hypothesis

Now P3: Millie is not disposed to drink the spoiled milk. In this section, we assume that P2—Millie wants to drink milk—is true, and argue that it's *possible* that P3 is also true. Suppose that you wanted to deny this possibility. Your claim would be that, given that Millie wants to drink milk, it *must* be

---

[14] This is a slight simplification. Braun suggests that Sarah may say more than one proposition in uttering (11).

[15] Now, if we were to stipulate that Millie does not want to drink milk—Braun makes such a stipulation in an analogous case in his §8.1—then she *should* be under pressure to retract. But that is not what's stipulated here in §4; rather, it's what's at issue.

that Millie is disposed to drink the spoiled milk. You'd then need a hypothesis about why Millie doesn't drink the spoiled milk, despite being disposed to drink it.

Here is such a hypothesis.

Start with something that everyone should agree on. How an agent acts depends not just on whether she has a certain desire and associated disposition, but also on what else she wants.[16] For example, suppose that Portia wants to buy a Porsche, and that she is disposed to buy a Porsche. She doesn't buy one, though, and that's because in addition to wanting to buy a Porsche, there's something else she wants: not to spend so much money that she is financially ruined. Her disposition to buy a Porsche isn't manifested because she wants this other thing.

According to the *other desires hypothesis* of Millie's inaction, Millie is like Portia. The hypothesis has two parts: (i) Millie *is* disposed to drink the spoiled milk, but (ii) she wants other things, preventing her disposition from manifesting.

Let's grant that Millie does want other things that bear on drinking the spoiled milk—e.g. she wants not to drink something sour, and she wants not to be sick to her stomach. The question is then whether her wanting these other things is interfering with the manifestation of a disposition to drink the spoiled milk—as the other desires hypothesis says. We think Millie has no such disposition.

To see why, contrast Millie with Portia, who, in being disposed to buy a Porsche, sees something in buying it: driving fast and making her friends envious. It makes sense that Portia would have a disposition to buy a Porsche—even though the disposition doesn't manifest itself—because a Porsche is alluring to her. But Millie sees nothing appealing at all in drinking the spoiled milk. What would the appeal even be? Everything that is normally appealing to Millie about milk is absent in the spoiled milk. Millie enjoys the mild flavor and smell of fresh milk; the spoiled milk is overpoweringly sour. Millie likes the smooth mouth feel of fresh milk; in the spoiled milk, the protein has separated from the whey, forming unpleasant clumps. Spoiled, separated milk doesn't even have the nice creamy look of fresh milk. Given that the spoiled milk has no appeal for Millie, why *would* she be disposed to drink it?

Even if you're not convinced by our argument against the other desires hypothesis in Millie's case, there are other cases relevantly like Millie's where the other desires hypothesis clearly fails. In these cases, the agent does not want any other things that could explain her inaction.

Consider Trina, whose neighbor has, much to Trina's dismay, just installed a full-scale plastic replica of Michelangelo's David. The sculpture is all too visible from Trina's kitchen window, and her view of it needs to be blocked tonight. Having a tree planted in between the sculpture and the window seems best: Trina wants to have a tree planted in her backyard by the end of the day. It so happens that Trina believes that the only trees available to her today are bonsais, which are too small to block her view of anything. Further, bonsais don't have the majestic look that Trina has always admired in trees of the size that could block the statue. Nothing that appeals to Trina about having a tree planted is present with a bonsai. The day ends without Trina trying to have a bonsai planted.

---

[16] Ashwell (2017) develops a theory on the interactions among desire-based dispositions.

The other desires hypothesis would say that (i) Trina is disposed to have a bonsai planted, but (ii) she wants other things, preventing this disposition from manifesting.

But we can easily suppose that Trina doesn't want any such things. Imagine that you go to Trina's backyard with a bonsai in hand, dig up a few inches of dirt, and tell Trina that you might plant the bonsai—how does she feel about it? Trina says that she doesn't care. As we know, nothing appeals to her about the bonsai. But neither is there anything unappealing. Having it planted comes at no cost to her. You are proposing to plant it for her, so she wouldn't have to get her hands dirty. And you wouldn't put the bonsai in a place that would stop Trina from planting a tree that could block the statue. Nor would you plant it in a place that would impede the route that she normally takes when she walks across her yard, or… Even if Trina did want not to get her hands dirty or to have her normal route unimpeded, her desires would have no impact on whether she has a bonsai planted.

As far as Trina is concerned, it's fine if the bonsai is planted, and fine if not. Trina is *indifferent*. There's nothing she wants either way about the bonsai. In particular, there's nothing that she wants about the bonsai that would prevent the manifestation of a disposition to plant a bonsai. This contradicts the other desires hypothesis.

Consider Portia for contrast again. Portia is *ambivalent*. She is at once both attracted to buying a Porsche (it would mean fast driving and envious friends) and repelled by it (she'd surely go bankrupt). The unappealing features of buying a Porsche overwhelm the attraction, which is why Portia does not buy a Porsche. The other desires hypothesis makes perfect sense of the situation. Given that Portia is both attracted to and repelled by the prospect of buying a Porsche, it's natural to think that she is both disposed *to* buy it, and that she wants other things that speak in favor of *not* buying it—things that prevent the disposition to buy it from manifesting. Not so with Trina. She is indifferent, neither attracted to nor repelled by the prospect of having a bonsai planted. It is her indifference that explains her inaction.

The other desires hypothesis fails with Trina. The point of the hypothesis is to explain why an agent does not act despite having a (hypothesized) disposition to act. No doubt Trina's case could be filled out so that Trina is disposed to have a bonsai planted, yet does not do so for some reason or other. But it clearly makes sense to fill it out in the way we have. If you want to maintain that Trina *must* be disposed to have a bonsai planted, you can't merely give a way of filling out the case so that Trina has an unmanifested disposition to have a bonsai planted; you must show that there is no possible way of filling it out as we have just done.

If you prefer Trina's case to Millie's, run our argument with Trina. Either way, P3 stands: the agent (Millie, Trina) is not disposed (to drink the spoiled milk, to have a bonsai planted).

6. In defense of the Satisfaction–Disposition Principle: on an "other-things-equal" clause

The final premise of our argument to defend is P1, which is an instance of the Satisfaction–Disposition Principle. We'll dispel one potential worry about the principle in this section and then others in §7 and §8.

When in a bold mood, philosophers state connections between desires and dispositions in the same form that we've stated the Satisfaction–Disposition Principle: if an agent is in such and such a desire state, then she is disposed to act thus-and-so-ly, given certain beliefs. When in a cautious mood,

philosophers add an "other things equal" clause: if an agent is in such and such a desire state, then, *other things equal*, she is disposed to act thus-and-so-ly, given certain beliefs.

You might worry that Millie's case calls for a cautious mood—that it calls for a version of the Satisfaction–Disposition Principle with an "other things equal" clause. If things were *un*equal with Millie, then our argument wouldn't go through.

Consider some ways for things to be unequal—ways for you to lack a disposition to do what you believe will satisfy your desire. You might be unaware of your desire, or have false second-order beliefs about your first-order beliefs about how to bring it about that your desire is satisfied, or be simply unable to bring it about that your desire is satisfied.

We can simply suppose that things are *not* unequal for Millie in these ways—that she is aware of her desires, that she believes that she believes that drinking the spoiled milk will bring it about that she drinks milk, and that she is perfectly able to drink the spoiled milk. Although there are many more ways for things to be unequal, we don't need to canvas them. Millie's case can be filled out so that things are not unequal in any of these additional ways. That's because her case, as already described, looks like a paradigm case where other things are equal. Everything is running smoothly: Millie isn't confused about her beliefs or desires, she's capable of drinking the spoiled milk, and the world is cooperating.

Using a version of the Satisfaction–Disposition Principle with an "other-things-equal" clause doesn't make a difference to our argument, since it makes perfect sense to think that other things are equal with Millie.

Zoom out for the moment and consider the broader dialectic. We have claimed that a certain case is possible, one where both P2 and P3 are true—where Millie wants to drink milk and is not disposed to drink the spoiled milk. Now we've added the supposition that other things are equal with Millie. But recall that for our argument to go through, we only need that there is *a* case where P2 and P3 are true and other things are equal. Our imagined interlocutor, on the other hand, must show that such a case (and all relevantly similar cases) is *impossible*.

7. In defense of the Satisfaction–Disposition Principle: on agent satisfaction vs. desire satisfaction

Another kind of worry about the Satisfaction–Disposition Principle doesn't concern the details of Millie's case, but rather the Satisfaction–Disposition Principle itself. You could grant the possibility of Millie's case as we've described it (that is, you could grant that it is possible that Millie wants to drink milk and is not disposed to drink the sour milk), yet deny that this shows anything about the satisfaction conditions of her desires. In this section we'll consider one objection to the Satisfaction–Disposition Principle; in the next section, another.

In arguing that desire satisfaction is *not* ways-specific (although they don't put it in those terms), Braun and Prinz distinguish desire satisfaction from what they call *agent satisfaction*. Desire satisfaction

is a matter of whether some one or other of an agent's individual desires is satisfied; agent satisfaction is a matter of whether the agent herself *feels* satisfied.[17]

With this distinction in mind, you might worry that the thesis that agents are disposed to satisfy their desires has been misunderstood: the thesis should *not* be understood in terms of individual desire satisfaction, (as it has been standardly (see e.g. §8 and the quotes on page 26)), but rather in terms of agent satisfaction. So the Satisfaction–Disposition Principle gets it wrong when it says that if you have *a* desire—an individual desire—that is satisfied in exactly the worlds where p is true, then you are disposed to what you believe will bring it about that p obtains. Rather, you are disposed to do what you believe will make yourself feel satisfied.

The worry is misguided. No doubt agents are in certain cases disposed to do what they believe will make themselves feel satisfied (although that doesn't mean they're not also disposed to do what they believe will satisfy their desires). But sometimes agents have desire-based dispositions that are *not* dispositions to do what they believe will make themselves feel satisfied. In such cases it's clear that desire satisfaction, not agent satisfaction, is what's at play.

Consider such a case: suppose that you want your name to live on after you die, and you do what you can to make it so. Suppose further that you don't in general feel good about merely *attempting* to reach your ends; rather, you feel satisfied only when you believe that your ends have been reached. (You're not one to hand out participation trophies.) As you work to make your name live on after your die—as you attempt to reach your end—you are unsure of whether you will succeed, and so you do not feel satisfied. And neither would you feel satisfied if you made your name live on after you die—if you in fact reached your end—since you don't feel anything at all after you die. You know all of this. So, as you do what you can to make your name live on, you neither experience nor anticipate any feeling of satisfaction.

You are disposed to do what you believe will make your name live on after you die. But your disposition is *not* to do what you believe will make yourself feel satisfied, since, again, you neither experience nor anticipate any feeling of satisfaction. Rather, your disposition is to do what you believe will satisfy one of your individual desires. The Satisfaction–Disposition Principle gets it right.

8. In defense of the Satisfaction–Disposition Principle: why accept it in the first place?

The final worry we'll consider about the Satisfaction–Disposition Principle is more general: why accept the Satisfaction–Disposition Principle in the first place?

The flat-footed answer is simple: the thesis that agents are disposed to satisfy their desires is true, and the Satisfaction–Disposition is a way of making this thesis precise. The subtler answer tells us why the Satisfaction–Disposition principle is a good way of making the thesis precise.

Recall how others have stated the thesis:

---

[17] Unlike Prinz, who identifies agent satisfaction with an agent feeling satisfied, Braun does not explicitly say what he means by "agent satisfaction." We read him as having the same thing in mind as Prinz. Fara (2003), Persson (2005, ch. 10), and Lycan (2012) also discuss something like this distinction.

> [T]he primitive sign of having a desire is trying to satisfy it. (Humberstone (, p. 107), riffing on Anscombe)
>
> [T]he actions a desire is a disposition to perform are those that would satisfy that desire provided the agent's operative beliefs were true. (Stampe, 1994, p. 246)
>
> [A] desire is manifested in…behaviour aimed at satisfying the desire. (Hyman, 2014, p. 85)

We can tease out two claims that are common among these quotes. The first is that from each desire, we can infer a disposition (or a trying, in Humberstone's case). The second is that this disposition is connected to the agent's desire in a certain way—it is a disposition to satisfy the desire. The Satisfaction–Disposition Principle, restated below, exemplifies both claims. It also allows us to make concrete predictions in a given case about whether an agent is disposed to do a certain thing, given her desires—something the above formulations don't allow us to do.

> *Satisfaction–Disposition Principle*
> If A has a desire that is satisfied in exactly the worlds where p is true, then A is disposed to do what she believes will bring it about that p obtains.

The crucial thing to establish is why this principle, and not some nearby principle, gets the connection between desires and dispositions right. Why would it be that it is exactly—i.e. all and only—the worlds where the desire is satisfied that matter to the disposition to satisfy it? Imagine that the principle were different.

Imagine, for example, that the principle were this: if A has a desire that is satisfied in *only* (but not necessarily all) worlds where p is true, then A is disposed to do what she believes will bring it about that p obtains. Then we would have a problem of disjunction introduction. Suppose Millie has a desire that is satisfied in exactly the worlds where she drinks fresh milk. She thereby has a desire that is satisfied only in worlds where she drinks fresh milk *or sprains her ankle*. She is not, though, disposed to do what she believes will bring it about that she drinks fresh milk or sprains her ankle.

Alternatively, imagine that the principle were this: if A has a desire that is satisfied in *all* (but not necessarily only) worlds where p is true, then A is disposed to do what she believes will bring it about that p obtains. Then we would have a problem of conjunction introduction. Suppose that Millie has a desire that is satisfied in exactly the worlds where she drinks fresh milk. She thereby has a desire that is satisfied in all worlds where she drinks fresh milk *and poisons her mother*. But Millie is not disposed to do what she believes will bring it about that she drinks fresh milk and poisons her mother.

The Satisfaction–Disposition Principle avoids both of these problems. Does it follow from the principle that Millie is disposed to do what she believes will bring it about that she drinks spoiled milk or sprains her ankle? No, because she does not have a desire that is satisfied in exactly the worlds where she does. Does it follow from the principle that Millie is disposed to do what she believes will bring it about that she drinks spoiled milk and poisons her mother? No, because she does not have a desire that is satisfied in exactly the worlds where she does.

9. Upshots: the dispositional role of desire satisfaction, revisited

We now have the premises, and so the conclusion: desire satisfaction is ways-specific. An agent may want p without having a desire that is satisfied in exactly the worlds where p is true.

This is a welcome conclusion: the thesis that desire satisfaction is ways-specific *explains* why agents are disposed to act as they are. Millie is not disposed to drink the spoiled milk *because* she is disposed to satisfy her desires and *she does not have a desire that is satisfied when she drinks the spoiled milk*. She has a desire-based disposition that is specific to certain ways of its being the case that she drinks milk because she has a desire whose satisfaction conditions are specific to certain ways of its being the case that she drinks milk. More generally, agents have ways-specific desire-based dispositions *because* they are disposed to satisfy their desires and *desire satisfaction is ways-specific*. (This prompts a question for the defender of the Satisfaction-is-Truth Principle: if desire satisfaction were *not* ways-specific, why would our desire-based dispositions be ways-specific, given that we're disposed to satisfy our desires?)

In addition to leading us to the conclusion that desire satisfaction is ways-specific, our argument gives us a new perspective on the dispositional role of desire satisfaction.

Consider, for example, that the following canonical principle connecting wanting and dispositions is false:

> *Want–Disposition Principle*
> If A wants p, then A is disposed to do what she believes will bring it about that p obtains.[18]

Millie wants to drink milk, but she not disposed to drink the spoiled milk—not disposed to do what she believes will bring it about that she drinks milk. Millie has a ways-specific desire-based disposition, which the Want–Disposition–Principle says is impossible. Recall that if an agent has a ways-specific desire-based disposition, then for some p, (i) she wants p; (ii) there are ways for p to obtain that she is disposed to bring about; but (iii) there are other ways for p to obtain that she is not disposed to bring about, *even if she believes that she only can bring it about that p obtains in those ways*. If an agent has a ways-specific desire-based disposition, then the antecedent of the Want–Disposition Principle may be true of her, but the consequent not.

The Want–Disposition Principle is false, but in it is a kernel of truth. To see the kernel, consider that the Want–Disposition Principle is entailed by the conjunction of the Satisfaction-is-Truth Principle and the Satisfaction–Disposition Principle, repeated here.

> *Satisfaction-is-Truth Principle*
> If A wants p, then A has a desire that is satisfied in exactly the worlds where p is true.

> *Satisfaction–Disposition Principle*
> If A has a desire that is satisfied in exactly the worlds where p is true, then A is disposed to do what she believes will bring it about that p obtains.

---

[18] Audi (1973, p. 4), Davidson (1976, p. 243), and Stalnaker (1984, p. 15), among many others, advocate principles in this spirit.

Think of the Want–Disposition Principle as factored into these two principles that entail it. Once we remove the false part, the Satisfaction-is-Truth Principle, we are left with the kernel of truth, the Satisfaction–Disposition Principle. Agents are disposed to satisfy their desires.

Another flaw in the Want–Disposition Principle sheds further light on the dispositional role of desire satisfaction. If the Want–Disposition Principle were true (and remember, we don't think that it is), we should be able to determine, just on the basis of certain of an agent's beliefs and whether she wants p, whether she is disposed to bring it about that p obtains in some certain way. But we can't do this. If all we know about Millie is that she wants to drink milk and that she believes that the only milk that's available to her is the spoiled milk, we can't determine whether she's disposed to drink the spoiled milk. What we need to know is whether drinking the spoiled milk is a way for her desire to be satisfied. Only then will we be able to pin down Millie's disposition.

10. Upshots: wanting, desires, and the Fara–Braun debate

Readers familiar with the debate between Fara and Braun may wonder how our argument relates to the locus of that debate: a set of three principles on which Fara and Braun disagree. The first principle is a version of the influential Relational Analysis of attitude ascriptions (e.g. Stalnaker (1988), Schiffer (2003)) as applied to desire ascriptions. The second two concern wanting, desires, and how they're related to each other.[19]

First, some terminology. We assume that at the level of logical form, the complement of "want" denotes a proposition, a standard assumption among semanticists (see e.g. Heim (1992) and von Fintel (1999)).[20] Let 'p' range over terms that denote propositions; let 'p' range over the corresponding propositions (ignoring any context-dependence in p); let 'A' range over the names of agents; and let 'A' range over the corresponding agents.

In stating the principles ourselves, we diverge slightly from Fara (2013)—she states all three principles as biconditionals, but her objection just concerns the left-to-right directions,[21] which is how we state them (and why we call them weak).

---

[19] There is a further question about what the noun "desire" denotes—i.e. what desires are (as opposed to *wanting* or *desiring*). This question, discussed by e.g. Schroeder (2004) and Braun (2015), is, we believe, beyond the scope of our paper.

[20] This assumption is compatible with the thought that at the level of *surface form*, the complement of "want" may not seem to denote a proposition—contrast e.g. "Millie wants to drink milk" with "Millie believes that she will drink milk."

[21] Braun makes the same point about the one of the principles, the Weak Specification Component, which we state just below.

Weak Relational Analysis
If ⌜ A wants p ⌝ is true, then A stands in the relation denoted by "wants" to p.[22,23]

Weak Content Component
If A stands in the relation denoted by "wants" to p, then A has a desire with p as its content.[24]

Weak Specification Component
If A has a desire with p as its content, then A has a desire that is satisfied in exactly the worlds where p is true.

Fara rejects the conjunction of the principles; Braun accepts it.[25]

How do the three principles relate to what we've said? Their conjunction, plus the following overwhelmingly plausible quotation principle *entail the Satisfaction-is-Truth Principle*.

Quotation
If A wants p, then ⌜ A wants p ⌝ is true.[26]

We repeat the Satisfaction-is-Truth Principe again for reference:

Satisfaction-is-Truth Principle
If A wants p, then A has a desire that is satisfied in exactly the worlds where p is true.

We accept Quotation and thus side with Fara in rejecting the conjunction of the three principles.

Though we reject the conjunction of these principles, our argument is silent on which principle or principles should be rejected (our argument is compatible with rejecting any given one or combination of them). Determining which should be rejected requires settling broader questions in the philosophy of language and philosophy of mind, questions beyond the scope of this paper. We will, however, suggest a way to proceed.

---

[22] Stated more precisely, the principle is as follows. For all A, A, p, and p: if A denotes A and p denotes p, then if ⌜ A wants p ⌝ is true, then A stands in the relation denoted by "wants" to p.

[23] Fara (2013) gives an instance of the principle: "'Lora wants Rudy to be in London' is true just in case Lora bears the relation expressed by "wants" to the proposition that Rudy is in London" (p. 250). Braun states the principle as follows: "If N is a proper name and S an infinitival phrase (with or without explicit subject), then ⌜ N wants S ⌝ is true iff the referent of N bears the relation expressed by "wants" to the proposition that S semantically expresses" (p. 144).

[24] For this principle and the next, see Fara's (2013) p. 253.

[25] More accurately, Braun accepts the latter two principles in conjunction with a different statement of the Weak Relational Analysis (see footnote 23).

[26] Stated more precisely, the principle is as follows. For all A, A, p, and p: if A denotes A and p denotes p, then if A wants p, then ⌜ A wants p ⌝ is true.

Each principle links a certain fact about wanting, desires, or desire ascriptions to another. The Weak Relational Analysis, for example, links the proposition denoted by the complement of "wants" with a proposition to which the agent stands in the relation denoted by "wants". In particular, it says that the proposition denoted by the complement of a "wants" ascription *is* a proposition to which the agent stands in the relation denoted by "wants". The Weak Content Component similarly says that the proposition to which the agent stands in the relation denoted by "wants" *is* a proposition which is the content of one of the agent's desires. In turn, the Weak Specification Component says that the truth conditions of the proposition that is the content of the agent's desire *are* the satisfaction conditions of the agent's desires. All of the principles link various facts about wanting, desires, and desire ascriptions by saying that the propositions that figure in these facts are identical.

Our argument shows, though, that not all of these propositions can be identical. "Millie wants to drink milk" is true, but Millie does not have a desire that is satisfied in exactly the worlds where she drinks milk. "Millie wants to drink milk" is true but the truth conditions of the proposition denoted by the complement of "want"—the proposition that Millie drinks milk—*are not identical to* the satisfaction conditions of any of Millie's desires. Rather, the relevant one of Millie's desires has satisfaction conditions that are *more specific* than this. That is to say, the satisfaction conditions of that desire are identical to the truth conditions of some proposition—perhaps the proposition that Millie drinks fresh milk—that *entails* the proposition that Millie drinks milk. Millie does not have a desire that is satisfied in exactly the worlds where she drinks milk, but she does (say) have a desire that is satisfied in exactly the worlds where she drinks fresh milk. Millie has a desire whose satisfaction conditions are ways-specific.

What we know, then, is that in attempting to link wanting, desires and desire ascriptions, at least one of the principles *underspecifies*—to use Fara's term—at least one of the relevant propositions. For example, it could be the Weak Content Component that goes wrong in this way. Then the proposition that is the content of the agent's relevant desire is more specific than the relevant proposition to which the agent stands in the relation denoted by "wants". If this is the case, we would propose replacing the Weak Content Component with the following principle: if A stands in the relation denoted by "wants" to p, then, *for some proposition* q *that entails* p, A has a desire with q as its content.[27] Here, the proposition that is the content of the relevant one of the agent's desires is not identical to the relevant proposition (p) to which she stands in the relation denoted by "wants". Rather, it is a more specific proposition (q). It needn't be, of course, that the problem is with the Weak Component Component. One of the other two principles could be the culprit instead. In that case, we would propose to replace those principles with alternatives that capture the specificity of the relevant propositions.

---

[27] Fara (2003, p. 159) advocates a similar principle: "A desire (or related attitude) ascription of the form 'A wants C' is true just in case A has a desire (or hope, etc.) with proposition Q as its exact content for some Q that entails the proposition expressed by the embedded clause C." (For a related view, see what call the "Quine-Hintikka" analysis of "want" ascriptions.) We believe that this is on the right track, but it's incorrect as it stands. It wrongly predicts that if ⌜ A wants q ⌝ is true, and q entails p, then ⌜ A wants p ⌝ is true. For example, it wrongly predicts that "I want to die quickly" entails "I want to die" (the example is from).

11. Conclusion

Our argument has been this: agents are disposed to satisfy their desires; desire-based dispositions are ways-specific; so, desire satisfaction is ways-specific. The Satisfaction-is-Truth Principle, which entails that desire satisfaction is *not* ways-specific, is false. In reaching this conclusion, we sidestep concerns about the probative value of intuitions about when people get what they want—intuitions on which Fara and Lycan rely—appealing instead to principles concerning the relation between desires and dispositions to act.

Our argument opens up certain questions. Satisfaction is not truth, so what is it? Desire satisfaction is ways-specific, but to which ways? We must reject one of the three principles at issue in the debate between Fara and Braun, but which? Finally, is the satisfaction of other attitudes—hoping, dreaming, fearing—also ways-specific? We've given a template for how to answer: look first to the attitude's dispositional role, and then work your way back to satisfaction.

Whatever the answers to these questions are, our argument shows that there's an important disanalogy between desire and belief. The Truth-is-Truth Principle is true but the Satisfaction-is-Truth Principle is false. Desire satisfaction is not truth by another name.

References

Anand P, Hacquard V. (2013). "Epistemics and Attitudes." *Semantics and Pragmatics* 6(8): 1–59.

Ashwell L. (2017). "Conflicts of Desire: Dispositions and the Metaphysics of Mind." In: Jacobs J. (ed), *Causal Powers*, Oxford University Press.

Audi R. (1973). "The Concept of Wanting." *Philosophical Studies* 24(1): 1–21.

Bach K. (1994). "Conversational Impliciture." *Mind and Language* 9(2): 124–162.

Bach K. (1997). "Do Belief Reports Report Beliefs?" *Pacific Philosophical Quarterly* 78(3): 215–241

Bach K. (2000) "Quantification, Qualification and Context: a Reply to Stanley and Szabo." *Mind and Language* 15(2&3): 262–283.

Bach K. (2001). "You Don't Say." *Synthese* 128(1):15–44.

Bach K. (2005). "Context Ex Machina." In: Szabo ZG (ed), *Semantics Versus Pragmatics*, Oxford University Press, pp. 15–44.

Braun D. (2011). "Implicating Questions." *Mind and Language* 26(5): 574–595.

Braun D. (2015). "Desiring, Desires, and Desire Ascriptions." *Philosophical Studies* 172(1): 141–162.

Condoravdi C., Lauer S. (2016). "Anankastic Conditionals are Just Conditionals." *Semantics & Pragmatics* 9(8): 1–2.

Davidson D. (1976). "Hempel on Explaining Action." *Erkenntnis* 10(3): 239–253.

Fara D.G. (2003). "Desires, Scope, and Tense." *Philosophical Perspectives* 17(1): 141–163.

Fara D.G. (2013). "Specifying Desires." *Nous* 47(2): 250–272.

von Fintel K. (1999). "NPI Licensing, Strawson Entailment, and Context Dependency." *Journal of Semantics* 16(2): 97–148.

Heathwood C. (2006). "Desire Satisfactionism and Hedonism." *Philosophical Studies* 128(3): 539–563.

Heim I. (1992). "Presupposition Projection and the Semantics of Attitude Verbs." *Journal of Semantics* 9(3): 183–221.

Humberstone I.L. (1990). "Wanting, Getting, Having." *Philosophical Papers* 99 (August): 99–118.

Hyman J. (2014). "Desires, Dispositions and Deviant Causal Chains." *Philosophy* 89(1): 83–112.

Lasersohn P. (1999). "Pragmatic Halos." *Language* 75(3): 522–551.

Lycan W.G. (2012). "Desire Considered as a Propositional Attitude." *Philosophical Perspectives* 26(1): 201–215.

Lycan W.G. (ms). "In what sense is desire a propositional attitude?" Unpublished manuscript; available at http://www.wlycan.com/uploads/8/0/5/1/80513032/desireprop.pdf.

Moltmann F. (2013). "Propositions, Attitudinal Objects, and the Distinction Between Actions and Products." *Canadian Journal of Philosophy* 43(5-6): 679–701.

Montague M. (2007). "Against Propositionalism." *Nous* 41(3): 503–518.

Persson I. (2005). *The Retreat of Reason: A Dilemma in the Philosophy of Life.* Oxford University Press.

Prinz J. (ms). "No Satisfaction? The Mundane Truth About Desires." Unpublished manuscript, as cited in Lycan (2012).

van Rooij R. (1999). "Some Analyses of Pro-attitudes." In: de Swart H (ed) *Logic, Game Theory and Social Choice*, Tilburg University Press, pp. 534–548.

Schiffer S.R. (2003). *The Things We Mean.* Oxford University Press.

Schroeder T. (2004). *Three Faces of Desire.* Oxford University Press.

Searle J.R. (1983). *Intentionality: An Essay in the Philosophy of Mind.* Cambridge University Press.

Soames S. (2005). "Quantification, Qualification and Context: A Reply to Stanley and Szabo." *Teorema: International Journal of Philosophy* 24(3): 7–30.

Soames S. (2008). "The Gap Between Meaning and Assertion: Why What We Literally Say Often Differs from What Our Words Literally Mean." In: Soames S. (ed), *Philosophical Essays, Volume 1: Natural Language: What It Means and How We Use It*, Princeton University Press, pp. 278–297.

Sperber D., Wilson D. (1985). "Loose Talk." *Proceedings of the Aristotelian Society* 86: 153–171.

Stalnaker R. (1984). *Inquiry.* Cambridge University Press.

Stalnaker R. (1988). "Belief Attribution and Context." In: Grimm R.H., Merrill D.D. (eds), *Contents of Thought*, University of Arizona Press, pp. 140–156.

Stampe D.W. (1986). "Defining Desire." In: Marks J. (ed), *The Ways of Desire*, Blackwell.

Stampe D.W. (1994). "Desire." In: Guttenplan S.D. (ed) *A Companion to the Philosophy of Mind*, Blackwell.

Strawson G. (1994). *Mental Reality*. MIT Press.

Whyte J.T. (1991). "The Normal Rewards of Success." *Analysis* 51:65–73

Yablo S. (2014). *Aboutness*. Princeton University Press.

Chapter 3

Moral Fetishism and Moral Worth

In Chapter 1, "What's Wrong With Wanting to Do What's Right?", I argued that the distinguishing feature of the moral fetishist is not that they have a desire to do what is right as such (as is often supposed) but that they have a desire to do what is right, *whatever rightness might be*. The moral fetishist's desire to do what is right is independent from any concern they may have for those features of actions that they take to relevant to what they ought to do—things like the welfare of others, kindness, keeping one's promises, or whatever else. The moral fetishist is therefore indifferent to the nature of right actions. The upshot is that although the good person and the moral fetishist may both have and be moved by a desire to do what is right, their desires have different contents.

In this chapter, I explore some of the implications of this account for how we should think about moral worth.

The implications of an account of moral fetishism for the question of what it takes for an action to have moral worth are necessarily limited; the absence of a morally fetishistic motivation does not imply the presence of a morally good one, so an account of moral fetishism cannot provide us with an account of moral worth. To say that an action is not motivated in a morally fetishistic way is not to say that it is thereby morally praiseworthy. (Even Kant's prudent grocer, who charges his customers fairly from an interest in preserving his reputation and his business, is not a moral fetishist). An account of moral fetishism may, however, give us reasons to reject certain widely held views of moral worth if those views entail that actions performed from morally fetishistic motives have moral worth. This is what my account of moral fetishism does; it gives us reasons to reject certain accounts of moral fetishism, while also placing constraints on what an adequate account of moral worth must look like.

On my account it is neither always morally fetishistic nor always morally praiseworthy to want to do what is right as such. It is perhaps unsurprising, then, that the accounts of moral worth for which my view makes trouble come from both sides of the most prominent divide in the literature on moral worth: the divide between those who hold that wanting to do what is right as such is incompatible with an action's having moral worth, and those who take it to be a necessary condition on moral worth. I will focus in particular on what I call "right-making reasons accounts" of moral worth (such as those offered by Nomy Arpaly and Julia Markovits), as well as a broadly Kantian account offered by Sliwa. The problem with these accounts, I will argue, is that they wrongly condemn certain morally praiseworthy actions as morally fetishistic, and count certain actions performed from morally fetishistic motives as having moral worth. These problems stem from a general failure to take into account the broader structure of the agent's desires and attend to differences in the content of superficially similar motivational states.

The plan for the paper is as follows. In section 1, I explain the theoretical role that the concept of moral worth is supposed to play. In section 2, I argue that considering the phenomenon of moral

fetishism gives us a reason to reject right-making reasons accounts of moral worth. In section 3, I argue that similar objections apply to at least some Kantian accounts of moral worth, such as that offered by Sliwa. In section 4, I conclude by drawing some broader conclusions about the nature of morally worthy action.

1. Moral worth

When it comes to questions of moral praise and blame, motives matter; it matters not merely which actions you perform, but the reasons for which you perform them. Someone who gives time or money to those in need does not deserve moral praise if their only motivation is to appear charitable, even though their action itself may be faultless.

Something similar is true of moral worth; it matters not merely that you do what is right, but that you do it for the right reasons. Just what the right reasons are is the central question in contemporary debates about moral worth.

But what do we mean, exactly, by "moral worth"? Though a person, a desire, or an intention might be appropriately regarded as morally praiseworthy, moral worth is a property of actions, not people or mental states. A right action has moral worth just in case the person who performs the action is motivated in the right sort of way (which, again, is the subject of much disagreement). Though the notions are distinct, moral worth is intimately related to moral praiseworthiness. Consider the following descriptions of our target concept:

> The moral worth of an action is the extent to which the agent deserves moral praise or blame for performing the action, the extent to which the action speaks well of the agent. (Arpaly, 2002, p.224)

> Morally worthy actions (the thought is) aren't just right actions—they are actions for which the agent who performs them *merits praise*. But not *all* praiseworthy actions have moral worth. We praise many actions for valuable or admirable qualities they have that are not moral—skillful actions, for example, are also praiseworthy. Morally worthy actions are ones that reflect well on the moral character of the person who performs them. (Markovits, 2010, p.203)

> Moral worth, as it is commonly understood, concerns whether (or the degree to which) an agent is praiseworthy for acting rightly (Isserow, 2019, p.252).

According to these views, a right action has moral worth iff the agent who performs the action is morally praiseworthy for performing it. Others take the notion of *moral credit* or *esteem* to be central:

> [Moral worth is] the special kind of value that a morally right action has when its rightness is creditable to its agent (Howard, 2021, p 157).

Still others argue that although moral praise or esteem may be appropriate responses to the moral worth of actions (and perhaps it is only morally worthy actions that are deserving of moral praise or esteem) they are not what an action's having moral worth consists in. Robert N. Johnson, for example, argues that "worthiness-of" interpretations of moral worth—interpretations of moral worth as worthiness of praise, esteem, or happiness-- do not comport with Kant's larger project of

uncovering the fundamental principle of morality. While looking to moral praiseworthiness and esteem may help us identify which actions have moral worth, moral worth is a matter of an action's exemplifying the fundamental principle of morality, not, ultimately, of its being morally praiseworthy.

Despite their differences, these descriptions should give us a general sense of the phenomenon we are interested in. And their differences won't matter much for my purposes. Each of these conceptions of moral worth maintain that morally worthy actions reflect a *good will*; there is something creditworthy, estimable, or morally praiseworthy about the motives of the person who performs morally worthy actions. And this is all I need to get my arguments off the ground. For if someone is motivated in a morally fetishistic way, and being so motivated is morally objectionable, then actions motivated in a morally fetishistic way do not have moral worth.

With a rough sense of the phenomenon of moral worth under our belt, we can now turn to a substantive account of the conditions on moral worth. As we've already noted, a morally worthy action must be motived in the right sort of way. A substantive account of moral worth gives us an answer to the question of just what the right sort of way is. Below I will consider two such accounts, what I'll call the *right-making reasons account* (developed by Markovits and Arpaly), as well as a Kantian account (one that requires the agent to have an explicitly moral motive) offered by Sliwa.

2. The right-making reasons account

The right-making reasons account of moral worth[1] claims that an action has moral worth if it is performed for the reasons *why* it is right—that is, if it is performed for the reasons that *make* it right.[2]

According to Markovits, for example,

> Morally worthy actions are those for which the reasons why they were performed (the reasons motivating them) and the reasons why they morally ought to have been performed (the reasons morally justifying them) coincide. (Markovits, 2010, p.230).[3]

And on Arpaly's view,

> [F]or an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons—that is, in response to the features that make it right (the *right reasons* clause).[4] (Arpaly, 2002, p.223).

---

[1] Properly speaking, the right-making reasons account of moral worth is not a single account, but a family of views. In what follows I'll often refer to "the" right-making reasons account for simplicity; my target is the core set of commitments all such accounts share.

[2] Versions of the right-making reasons account of moral worth have been given by Arpaly (2002, 2014, 2015) and Markovits (2010, 2012), among others.

[3] Markovits's version of the right-making reasons account is what she calls "The Coincident Reasons Thesis": "*My action is morally worthy if and only if my motivating reasons for acting coincide with the reasons morally justifying the action."* (italics in original). (2010, p.205)

[4] There are several differences between Markovits's and Arpaly's accounts of moral worth; perhaps the most significant is that Arpaly, but not Markovits, thinks that moral worth comes in degrees, where an action's degree of moral worth is determined by the agent's degree of concern for moral reasons.

According to these accounts, a person need not judge her action to be right, let alone be motivated by such a judgement, in order for that action to have moral worth. She simply needs to be motivated by the features of her situation that make that make it the right thing to do. The case of Mark Twain's Huck Finn, who helps his enslaved friend Jim escape recapture despite believing that he morally ought to return him to his "rightful owner", is the now-classic case used to illustrate and motivate this view. Despite believing that doing so is wrong, Huck helps Jim because he recognizes Jim's humanity[5]; despite both the racist ideology of his time and his own misguided moral beliefs, Huck is able to see and respond appropriately to the morally relevant features of his situation. As a result, Huck succeeds in doing what is right, and his actions have moral worth because he is motivated by the features of his action that make it right.

As the case of Huck Finn makes clear, good people can have false beliefs about what morality requires of them, but nonetheless manage to do the right thing for the right reasons. In some cases (like Huck's) the agent's false moral beliefs are the product of their cultural context. In others, such as Arpaly's case of a student who espouses Ayn Rand's views while going out of their way to help others, their false moral beliefs are attributable to the fact that they are "good people who happen to be incompetent abstract thinkers" (Arpaly 2002, p.230). Whatever the explanation for their failure to see the right action *as right*, the claim is that it is the quality of a person's motives – not the correctness of their moral beliefs – that determines whether their actions have moral worth. Moreover, a person's motives can be good (and more specifically, *moral worth-conferring*) even if the person does not recognize, in explicitly moral terms, what they ought to do. What matters for moral worth is that the person is appropriately responsive to moral reasons, not that they recognize them as such.

One appealing feature of the right-making reasons account of moral worth is that it is able to accommodate the fact that when a person's action has moral worth, it is not an *accident* that they did what was right.[6] Compare Huck's motives in helping Jim to the motives of Kant's prudent shopkeeper. In charging his customers fairly, Kant's shopkeeper does what is right, but the fact that he does what is right is explained by a lucky coincidence between what would promote his own selfish interests (to preserve his reputation, and therefore profits) and what morality requires. In contrast, Huck does what is right because he recognizes Jim's humanity—that Jim is not property to be bought and sold. Insofar as this is what makes helping Jim the right thing to do, it is no accident that Huck does what is right.

This feature of morally worthy actions— *non-accidentality*—is widely taken to provide a constraint on any account of moral worth.[7] The motivating thought here is that motives which only happen to align with the dictates of morality, but could easily have failed to do so, do not reflect a good will; given their motives, it is merely a matter of luck that the person stumbles upon doing what is right.

---

[5] At least on some interpretations of the story.

[6] For arguments that right-making reasons accounts do not in fact secure the kind of non-accidentality required for moral worth, see Singh, K. (2020) and Johnson King, Z. (2020).

[7] According to Johnson King (2020), non-accidentality is not merely a constraint on accounts of moral worth, it is the defining feature of morally worthy actions. This is important because, according to Johnson King, moral praiseworthiness and non-accidentality come apart; there are actions for which an agent may be morally praiseworthy, but where the connection between the agent's motives and their performance of the right action does not satisfy non-accidentality.

The right-making reasons account of moral worth satisfies non-accidentality, its proponents claim, because when an agent is motivated by right-making reasons her motives necessarily coincide with the requirements of morality. An action's rightness and the reasons in virtue of which it is right go hand-in-hand, so when a person is motivated by the latter it is no accident that they hit upon the former.

If we buy the claim that the actions of people like Huck Finn have moral worth, then we are also committed to the claim that wanting to do what is right as such, or *under that description*, is not necessary for an action to have moral worth. The right-making reasons account therefore stands in contrast to broadly Kantian approaches to the question of moral worth. For Kant, only actions performed from the motive of duty have moral worth. Acting from the motive of duty is an exercise of the will which requires acceptance of the principle upon which one acts. And this, Kant thought, requires a kind of reflective endorsement—a recognition of one's action *as right*. Yet this is precisely the kind of cognitive achievement that is out of reach of people like Huck Finn.

Cases like Huck's are supposed to undermine the necessity of being moved by a desire to do what is right as such for moral worth. But denying that a self-consciously moral motive is required for moral worth is consistent with thinking that such motives are sufficient. One might think that while you don't need to conceive of your actions and your reasons for acting in explicitly moral terms in order for them to have moral worth, there isn't anything *wrong* with (correctly) conceiving of them in moral terms. However, the right-making reasons account of moral worth is committed to denying this attractive thought, because it entails that an action cannot have moral worth if it is motivated by the desire to do what is right as such.

To see why, consider again the central claim of the right-making reasons account: an action has moral worth only if the agent is motivated in the performance of the action by the reasons that make it right. But the fact that an action is right isn't what *makes* it right; actions are made right by things like the fact that the action is fair, or that it best respects the rights of all those affected, or that it contributes more to wellbeing than other available actions. Rightness itself doesn't make a contribution. It follows that actions motivated by a desire to do what is right as such do not have moral worth. This is not a new observation; it is a result that many of those who defend right-making reasons accounts of moral worth embrace.[8]

The right-making reasons account entails that the fact that an action is right is not what motivates those whose actions have moral worth. But in the absence of some independent reason to find this implication plausible, we might take it to be a strike against the view. After all, is it really so obvious that there is something wrong with wanting to do what is right as such?

[8] Markovits (2010), for example. Note that Markovits allows exceptions to this claim in cases where one is not in a position to know what makes a given action right, but nonetheless knows (for example, by moral testimony) that it is right. On her view, what matters for moral worth is what a person subjectively ought to do; what they ought to do given their evidence. The relevant right-making reasons, then, are subjective right-making reasons— reasons that, given the agent's evidence, make it the case that she subjectively ought to perform some action. Moral testimony to the effect that some action is right can be what makes that action subjectively right. A person who is motivated to perform some action upon learning that it is right may be well-motivated, then, even if they are motivated by the fact that the action is right.

Admittedly, the answer is "yes" in some cases. Unsurprisingly, these are the cases that proponents of the right-making reasons account tend to focus on. Some of these involve people whose motives look objectionably impersonal, as in Williams' (1981) case of a man who saves his drowning wife rather than a stranger, because *it is my wife, and in situations like this it is morally permissible to save one's wife,* or Stocker's (1976) case of a friend who visits you in hospital not because she cares about you, but because she believes it is her moral duty. In each of these cases, although agent does what is right, their motives strike us as wrong. They are motivated by concerns about the rightness of actions, but independently of any concern for what makes their actions right. These are examples of "moral fetishism".

Other problematic cases involve people who have mistaken moral beliefs and, as a result, end up doing what is wrong despite having a desire to do what is right as such. Arpaly asks us to imagine a Nazi who wants to do what is right as such, but who believes that right actions are those that bring "glory for the Aryan race and the destruction of the Jewish people." Obviously, this person's motives are morally contemptible, and no action those motives might produce would have moral worth. As Arpaly correctly points out, the fact that the Nazi wants to do what is right does nothing to redeem them or their actions, "not even a little bit"(Arpaly, 2014, p.63).

And yet, in other cases, wanting to do what is right as such seems perfectly natural and does not appear to reflect poorly on a person's motives or character. As I argued in chapter 1, many cases of moral uncertainty fit this mold; a person is uncertain how they ought to weigh a variety of moral considerations and as a result they don't know what they ought to do. Despite this, they want to do what is right because they believe that the right thing to do *just is* whatever is favored by the balance of those considerations. Moreover, those considerations matter to them— they care about doing what is fair, or what respects the rights of others, or what best promotes wellbeing (and so on for the various morally relevant considerations that might be in play in any given case). It is a stretch to describe such people as moral fetishists. Moreover, it is certainly not obvious that if they go on to perform the right actions, their actions do not have moral worth.

The right-making reasons account claims that actions performed from a desire to do what is right as such *never* have moral worth, but this seems to fly in the face of our intuitions in many cases. This may be a problem particularly for those like Arpaly and Markovits who take the moral worth to be a kind of moral praiseworthiness; for surely in many of these cases the agent's action seems morally praiseworthy (as does the agent's concern for the rightness of her actions).

While I think this has the makings of a powerful objection to the right-making reasons account, I won't develop it further here. The point is just that these cases put some pressure on the right-making reasons account of moral worth. All other things equal, it would be better to have a view of moral worth that doesn't exclude them by fiat. In any case, the argument I will develop in the rest of this section does not rely on the claim that actions performed out of a desire to do what is right as such can have moral worth.[9] Instead, I will press what I take to be a more decisive objection against right-making reasons accounts: they incorrectly classify some clear cases of moral fetishism as cases of morally worthy action.

[9] In "What's Wrong With Wanting to Do What's Right?" I argued for a weaker claim: that such actions are not motivated in a morally fetishistic way.

The objection stems from the fact that the charge of moral fetishism can be appropriate even when the target of the criticism does not have a desire to do what is right as such. Consider someone who has a single-minded devotion to doing what's fair, *whatever that might be*. Their concern with fairness is independent of any of those things that make fairness morally valuable: if fairness requires people getting what they deserve, so be it; if it requires punishing people mercilessly, so be it. Though they are committed to fairness, their commitment to fairness does not depend on what acting fairly actually turns out to involve. We can think of this person as a *fairness fetishist*.

When the fairness fetishist has true beliefs about which actions are fair, they will succeed in doing what is actually fair. And in many (if not all) of those cases, their action will thereby be morally right. So they do what is right, and they do it for the reasons that make it right. By the lights of the right-making reasons account, their actions have moral worth. But intuitively, the motives of the fairness fetishist are no better than those of the moral fetishist who is indifferent to the nature of right action. Unlike the person who fetishizes moral rightness, the fairness fetishist acts for right-making reasons, but their motivations are no less objectionable.

In fact, it seems that for each kind of reason that might make an action right, we can imagine someone who has a fetishistic concern with doing things for that reason. (Imagine, for example, someone who is fetishistically concerned with doing only what is lawful, or what is selfless.) The right-making reasons account tries to explain why paradigmatic cases of moral fetishism are not cases of moral worth (the fact that an action is right is not what makes it right), but it does not have the resources to exclude the actions of those who are moved by these other, lower-order varieties of moral fetishism.

3. Sliwa's Kantian account of moral worth

In the previous section, I argued that right-making reasons accounts misclassify some cases of moral fetishism as cases of morally worthy action. In this section, I will argue that a similar objection applies to a recent version of the Kantian account of moral worth proposed by Sliwa. According to Sliwa's account,

> A morally right action has moral worth if and only if it is motivated by concern for doing what's right (conative requirement) and by knowledge that it is the right thing to do (knowledge requirement). (Sliwa, 2015b)

What makes this a "Kantian" account of moral worth is that – in sharp contrast to right making reasons accounts – it holds that morally worthy actions *must* be motivated by a desire to do what is right as such. But like Arpaly, Sliwa recognizes that wanting to do what is right as such is consistent with having deeply mistaken views about what rightness actually requires in particular cases. (Recall Arpaly's case of the Nazi who wants to do what is right as such.) This is what leads Sliwa to require that in addition to wanting to do what is right, the agent must *know* what is right.

Sliwa's knowledge requirement thus plays a similar role to Arpaly and Markovits's requirement that agents act for right-making reasons: it aims to ensure that morally worthy actions are motivated by the kinds of things that actually matter from a moral perspective. In the previous section, we saw that Arpaly and Markovits's attempt to satisfy this requirement fails. That is, an agent can be motivated to act by the reasons that make her action right, but nonetheless have objectionably fetishistic motives. In this section, I will develop a similar objection to Sliwa's account of moral

worth: an agent can fulfill both Sliwa's cognitive and knowledge requirements on moral worth while nonetheless having entirely fetishistic motives.

In chapter 1, I argued that wanting to do what is right as such is only morally fetishistic in some cases. What makes such a desire morally fetishistic, I argued, is that it is not appropriately conditional on what rightness turns out to require. The moral fetishist wants to do what is right regardless of what rightness turns out to be, whereas the virtuous agent wants to do what is right only on the condition that what rightness requires is not too far removed from her substantive conception of morality. When the good person wants to do what is right as such (in cases of moral uncertainty, for example), she wants to do what is right as such because she believes that the right thing to do *just is* whatever action is supported by the kinds of things that she independently cares about, such as the wellbeing of others, treating people fairly, and so on.

If this account of moral fetishism is correct, then there is no obvious reason why a person with a morally fetishistic desire to do what is right could not sastify Sliwa's two requirements. Suppose that John is firmly committed to doing what is right, and always knows just what that is, as well as why. As a result, John's actions always have moral worth by the lights of Sliwa's account, because he is invariably "motivated by concern for doing what's right … and by knowledge that it is the right thing to do." It may sound like I have just described a moral saint, but notice that nothing I have just said rules out the possibility that John's desire to do what is right is fetishistic in nature. That is, we can suppose that

(a) John wants to do what is right, knows what is right, and is motivated to do what he knows is right

while also supposing that

(b) John's desire to do what is right is entirely unconditional on what rightness turns out to require.

But if we make this additional supposition, then the claim that John's actions would nonetheless have moral worth is deeply implausible, because John would be a moral fetishist *par excellence*.

To see this, consider how John would have been motivated if, keeping his actual desires fixed, he had formed very different moral beliefs. That is, suppose that John comes to believe (incorrectly) that everything he thought he knew about morality is false. Previously, he believed that morality enjoins us to perform actions that promote the wellbeing of others, and treat them fairly, and so on; now he has come to believe that morality requires doing whatever would maximize only his own pleasure. As a result, his firm commitment to doing the right thing – the very same desire that once motivated him to act rightly on a consistent basis – now motivates him to spend all his money on spa treatments.

I think it is clear that acting out of a desire to do what is right that is unconditional in this way would not redound to John's credit, *even when he acts rightly*.[10] Moral worth requires more than knowing what

[10] To be clear, I am not assuming that a person's counterfactual motives or desires make a difference, in their own right, to the moral worth of their actions. I am inclined to think that is only a person's actual motives—those motives that in fact moved them to do what is right—that matter for thinking about the moral worth of

is right, and being motivated to act in accordance with that knowledge (if indeed it requires those things at all). It requires, further, that the agent care about the right kinds of things–such as the interests of other moral agents–and that her actions be motivated by concern for those things. But John's case shows that an action can satisfy Sliwa's account of moral worth without being motivated by that sort of concern. This in turn means that Sliwa's account of moral worth is at least incomplete: some further requirement is needed to avoid misclassifying morally fetishistic actions like John's as having moral worth.

4. Conclusion

I have argued that two popular accounts of moral worth–the right-making reasons account of moral worth and Sliwa's Kantian account–misclassify some morally fetishistic actions as having moral worth. In closing, I want to consider where those arguments leave us. What more general lessons can we draw about the nature of moral worth?

As we have seen, it is an important constraint on any account of moral worth that it capture the fact that morally worthy actions are not morally right by accident. That is, given the agent's motives, it is not a matter of luck that they do what is right. The right making reasons account attempts to capture this non-accidentality by requiring that actions be performed for the reasons that make them right. Sliwa's Kantian account attempts to capture it by requiring that the agent *know* what the right thing to do is, and be motivated by this knowledge.

The arguments of the previous two sections show that, insofar as each account misclassifies as morally worthy actions that are motivated by morally fetishistic desires, neither account entirely succeeds in capturing the sort of non-accidental connection between an agent's motives and actions that is required of moral worth. One implication of the foregoing discussion is that having a good will is at least partly a matter of *wanting* and *caring about* the right sorts of things. A person that does the right thing out of an unconditional desire to do what is right as such might reliably do the right thing–but only provided that they do not change their mind about what morality is all about. Similarly, an agent might do the right thing for the reasons that make their action right, but only be disposed to treat the fact that an action would be fair as a reason (for example) because they have fetishized the value of fairness, rather than because they are concerned about others and want to give them their due. In each of these kinds of cases a person could, consistent with continuing to have *exactly the same desires*, come to be motivated to perform truly horrific actions for truly horrific reasons. These desires do not, therefore, reflect a good will, and actions motivated by them do not have moral worth.

When the good person cares about doing the right thing, or doing what is fair, by contrast, she cares about doing the right or fair thing because she believes that the right or fair action *just is* the action that achieves the balance of the things she believes to be relevant to rightness or fairness, and that she independently cares about. And it is the fact that she cares about the right things–the things that really do matter from a moral perspective–that make it no accident that she does the right thing.

---

the action performed. (This allows for the possibility that a person might perform an action with moral worth despite acting out of character.) However, knowing how a person is disposed to act under counterfactual conditions in which they form different moral beliefs often tells us something about the *content* of their actual motives, and so is relevant to assessing the praiseworthiness of those motives.

An adequate account of moral worth, then, will need additional constraints – constraints not found in either of the accounts considered here – to avoid classifying morally festishistic actions as having moral worth. In particular, it will need to place further constraints on the content of the desires that ultimately explain why the agent acts as she does. While a detailed account of those constraints is beyond the scope of this paper, the general shape that they will need to take is relatively clear. The agent's motivations for doing what is right must ultimately be grounded in an independent concern for the right sorts of things – a concern that would naturally lead them to recoil from warped conceptions of what morality requires, and so provide a degree of immunity against acting in accordance with such a conception.

What exactly are the "right sort of things," though? So far, my conclusions on this front have mostly been negative. An abstract, fetishistic concern for "morality" does not count. Neither does a similarly abstract concern for fairness, beneficence, or any other explicitly moral value. It seems to be a matter about being genuinely, nonderivatively concerned for others—but spelling this thought out will have to wait.

References

Arpaly, N. (2002). Moral Worth. *The Journal of Philosophy,* 99(5), 223-245.

Arpaly, N. (2014). Duty Desire and the Good Person: Towards a Non-Aristotelian Account of Virtue. *Philosophical Perspectives,* 28, 59-74.

Arpaly, N. (2015). Huckleberry Finn Revisited: Inverse Akrasia and Moral Ignorance. In Clark, R., McKenna, M., and Smith, A. M., editors, *The Nature of Moral Responsibility,* Oxford University Press.

Howard, N. R. (2021). The Goals of Moral Worth. *Oxford Studies in Metaethics,* 16.

Isserow, J. (2019). Moral Worth and Doing the Right Thing by Accident. *Australasian Journal of Philosophy,* 97(2), 251-264.

Johnson, R. N. (2009). Good Will and the Moral Worth of Acting from Duty. In Hill, T. E. Jr., editor, *The Blackwell Guide to Kant's Ethics.* Blackwell.

Johnson King, Z. A. (2020). Accidentally Doing the Right Thing. *Philosophy and Phenomenological Research*, 100(1), 186-206.

Markovits, J. (2010). Acting for the Right Reasons. *Philosophical Review,* 119(2), 201-242.

Singh, K. (2020). Moral Worth, Credit and Non-Accidentality. *Oxford Studies in Normative Ethics*;

Sliwa, P. (2015a). Praise without Perfection: A Dilemma for Right-Making Reasons. *American Philosophical Quarterly,* 52(2).

Sliwa, P. (2015b). Moral Worth and Moral Knowledge. *Philosophy and Phenomenological Research,* 93(2), 393-418.

Stocker, M. (1976). The Schizophrenia of Modern Ethical Theories. *The Journal of Philosophy,* 73(14), 453-466.

Williams, B. (1981). *Moral Luck*. Cambridge: Cambridge University Press.