

# The Puzzle of Pure Moral Motivation

\*Draft\*

8/9/2016

Adam Lerner

Dear CHillMeta readers: This is a long paper. For those of you who don't have enough time to read the whole thing, here is some guidance that should help you choose your own adventure. If you are interested in arguments for the view that it is rational to have a *de dicto* desire to do what is right, you should read section 1. If you are willing to assume this view for the sake of argument and are more interested in seeing what follows for metaethics, you can skip section 1. If you want to know why the view poses a problem for non-cognitivism, read section 2. If you want to know why it poses a problem for Cornell Realism, read sections 3 and 4. If you want to know why it poses a problem for non-naturalism, read sections 3 and 5. If you want to know why it poses a problem for analytic naturalism, read section 6.

Comments very welcome: [adamjl@princeton.edu](mailto:adamjl@princeton.edu)

## Introduction

Some people care about doing the right thing *as such*. They want to do the right thing *under that description*. These people have what has come to be known as a *de dicto* desire to do what is right.<sup>1</sup>

Some people who have a desire to do what is right *as such* desire to do what is right only as a means. They want to do the right thing in order to avoid punishment, or as a way of doing what God or their mother would approve of, or because it would be a sign that they are predestined for salvation. But many care about doing the right thing for its own sake: because it is the right thing. They have a *final* desire to do what is right *as such*. This desire is an instance of what I will call *pure moral motivation* (PMM).<sup>2</sup> Other instances of PMM include the desire to avoid doing what is wrong *as such*, the desire to promote what is good *as such*, and the desire to prevent what is bad *as such*.

---

<sup>1</sup> The terminology originates with Smith (1994), p. 74-76.

<sup>2</sup> McGrath (2015) provides the inspiration for both the terminology and the paper title.

Contemporary views in metaethics have been built to satisfy a number of familiar desiderata: accommodating Moorean open questions, explaining the connection between moral judgment and moral motivation, accounting for the supervenience of the moral on the non-moral, and so on. In this paper, I introduce a new desideratum: vindicating the rationality of PMM. I argue that PMM is rational in some important cases, and that any adequate view must accommodate this fact. The puzzle of PMM is to explain how that can be done. I argue that solving the puzzle poses a serious challenge for each of the standard views in metaethics.<sup>3</sup>

In section 1, I argue that PMM is at least sometimes rational. In section 2, I explain why non-cognitivism has trouble accommodating the rationality of PMM. In section 3, I introduce two principles regarding the conditions under which final desires and final preferences are rational. In sections 4 and 5, I draw on these principles to argue that synthetic naturalism and non-naturalism both have trouble accommodating the rationality of PMM. In section 6, I explain why analytic naturalism has trouble accommodating the rationality of PMM. Section 7 responds to an objection and section 8 concludes.

## Section 1: The Rationality of Pure Moral Motivation

At the heart of moral practice lies moral inquiry. And at the heart of moral inquiry lies the constitutive aim of discovering the moral truth; one cannot genuinely engage in moral inquiry without having a desire to discover the moral truth.

People regularly engage in moral inquiry. So they must have a desire to discover the moral truth. Why would anyone want to discover the moral truth?

Any number of things might motivate someone to try to discover the moral truth. They might get paid to do it. They might want to predict what virtuous people are likely to do in certain circumstances so that they may exploit their virtuous character. They might just be curious.

But this is not why most people want to discover the moral truth. Most people want to discover the moral truth because they want to act in accordance with the moral truth. They want to do the

---

<sup>3</sup> While I focus in this paper on specifically *moral* motivation, I believe the problem arises for pure normative motivation generally.

right thing *as such*.<sup>4</sup> And they want to do it for its own sake. They have a *final* desire to do the right thing *as such*.<sup>5</sup> They exhibit PMM.

In this section, I argue that people who engage in moral inquiry are often driven by PMM, and that this is significant. It is significant because it is often perfectly rational to engage in moral inquiry.<sup>6</sup> If it is often perfectly rational to engage in moral inquiry, and if perfectly rational moral

---

<sup>4</sup> In desiring to do the right thing *as such*, one conceives of the thing one wants to do *as* the right thing. One desires it under that description, *the thing that has the property of being right*. When one has a merely *de re* desire to do the right thing, one conceives of the thing one wants to do under some alternative guise—say, *as the thing that saves the drowning child*. A *de re* desire to do the right thing counts as a desire to do the right thing because the object of the desire *is* right, not because it is conceived to be.

<sup>5</sup> We can understand final desires by contrasting them with instrumental desires. S's desire that P is an instrumental desire just in case S has that desire only because (a) S desires something else Q, and (b) S believes that P raises the probability of Q either by causing, realizing, or signifying Q. (When I say that S has that desire *only because* of these further facts, I mean these further facts together constitute S's *sole rationale* for that desire.) S's desire is a final desire just in case S's rationale for their desire does not consist *only* in pairs of claims like (a) and (b).

To illustrate, consider Jonathan. Jonathan has a desire to do the right thing *as such* only because he has a distinct desire to do what will get people to like him and the belief that doing the right thing will cause people to like him. Jennifer has a desire to do the right thing *as such* only because she has a distinct desire to do what her mother wants her to do and the belief that her mother wants her to do what is right. Calvinist Jonah has a desire to do the right thing *as such* only because he has a distinct desire to be saved and a belief that doing the right thing is a sign that he has been saved. Jonathan's, Jennifer's, and Jonah's desires are all instrumental desires. But now consider George. George has a desire to do the right thing *as such*, but not because doing the right thing raises the probability that he'll get something else he wants. If asked why he wants to do the right thing, he would sincerely reply: "Because it's right." George's desire is a final desire. And as a final desire for the right thing *as such*, it is an instance of PMM.

<sup>6</sup> What am I saying when I say it is "rational" for people to have a motivation? For purposes of this paper, a motivation is rational if and only if the person who has the motivation possesses at least *some* good reason to have that motivation. And for someone to possess good reason to have a motivation, it is not enough that there be some reason for them to have it: they must possess that reason. It must be within their ken. (Cf. Lord, 2015)

inquiry is often driven by PMM, then the PMM that often lies behind moral inquiry must itself be perfectly rational. And if the PMM that lies behind moral inquiry is perfectly rational, then it is at least sometimes rational to have PMM.

I will make my argument by considering three different forms of moral inquiry: moral deliberation, gathering novel information, and moral deference.

## 1.1 Moral Deliberation

Delilah is the owner of a factory farm. She knows everything there is to know about what goes on inside of a factory farm. She also knows everything there is to know about how industrial agriculture impacts the environment, the economy, and people's gustatory experiences. But imagine Delilah was born into the trade and never thought about whether it might be wrong to raise animals on factory farms. Or maybe she has given it some thought, but she hasn't come to a settled view. She now faces a decision: should she sink more time and effort into trying to figure out whether running a factory farm is right?

Suppose Delilah chooses to make the investment. This seems like a perfectly good and rational decision to make.<sup>7</sup> Why would Delilah make this decision? Clearly, she has a desire to know what is morally right in this case. But *presumably* she doesn't want this bit of moral knowledge for its own sake. She wants it so that she may be in a position to do what is right. And she wants to do what is right. Or at the very least, she wants to avoid doing what is wrong.

Importantly, this desire to do what is right must be a desire to do what is right *as such*. If it were merely a *de re* desire to do what is right, we could not explain why Delilah chooses to deliberate. If Delilah merely had a *de re* desire to do what is right, then she would desire to do it under some *non-moral* description. But if she wanted to do what is right under some *non-moral* description, she would already know what act she would have to perform in order to satisfy her desire—after all, she already knows all of the morally relevant non-moral facts about the case.

---

<sup>7</sup> In thinking that Delilah exercises good moral agency in choosing to engage in moral inquiry, I agree with Jamie Dreier: “A good moral agent who believes utilitarianism but thinks that there is a chance that a rights-based theory is correct would surely want to investigate the possibility.” (2000, p. 634)

So when Delilah chooses to deliberate, she must be acting on a desire to do what is right *as such*.<sup>8</sup> But is her desire an instance of PMM? In order to count as an instance of PMM, it must be a final desire rather than an instrumental desire. Is it? Supposing Delilah is a virtuous agent, then it is a final desire.<sup>9</sup> She does not want to do the right thing in order to avoid punishment, or as a way of doing what her mother wants, or to confirm that she is going to heaven. She wants it for its own sake.<sup>10</sup>

Delilah acts on PMM when she chooses to deliberate. This shows that PMM is rational. Why? Because Delilah is perfectly rational in choosing to deliberate. And if an act is rational, then the desires that motivate it are rational. Since Delilah is driven by PMM, it follows that her PMM is rational.

---

<sup>8</sup> I am assuming here and in the rest of the paper that the Humean theory of motivation is true; a belief that doing the right thing *as such* is right could not alone explain why Delilah decides to deliberate. But my thesis does not rest on this assumption. If you are an anti-Humean, feel free to replace any instance of “being driven to A by a final desire to do the right thing *as such*” with “being driven to A by a belief that doing the right thing *as such* is right”. Likewise, you can interpret the epistemic constraints defended in section 3 as constraints not on when it is rational to finally desire P or finally prefer P to  $\sim$ P, but as constraints on when it is rational to be driven by a belief that P is right.

<sup>9</sup> I disagree here with Smith (1994), Arpaly (2002), and Weatherston (2014), all of whom hold that virtuous agents do not have final *de dicto* desires to do what is right.

<sup>10</sup> I claim in the text that pure moral inquiry of the sort that Delilah engages in could only be explained by a final *de dicto* desire to do what is right. Jamie Dreier argues that a second order final (*de dicto*) desire to have final desires to do whatever is right (*de re*) can also explain why people engage in moral inquiry: “If David thought it possible that moral side constraints were a fundamental feature of true morality, his second order desire would motivate him to investigate. For he wants to have desires of a certain sort (namely, ones whose objects are morally right actions), and he knows that there is a significant chance that the only way he can have desires of that sort is by thorough moral reflection and investigation.” (2000, p. 632) While this may be right, this second order final desire is still an instance of PMM. And as an instance of PMM, every reason for thinking that prominent views in metaethics cannot account for the rationality of a final *de dicto* desire to do what is right is also a reason for thinking that they cannot account for the rationality of this second order desire. For simplicity, then, I will only discuss the final *de dicto* desire to do what is right in what follows.

Delilah's case may sound strange. Is any actual person ever in Delilah's position? Does anyone know all of the relevant non-moral facts, or at least take themselves to know all of the relevant non-moral facts, and still engage in moral deliberation?

Indeed, it is a pervasive feature of ordinary people's moral lives. When people find themselves puzzling over whether it is okay to get an abortion, whether they are giving enough to charity, or whether they should stop eating meat, they are not just wondering about how the world is in non-moral respects. They often take themselves to know all of the non-moral facts. Likewise, when normative ethicists puzzle over the correct solution to the non-identity problem, or whether it is right to save one person from death or three people from paralysis, they are almost never wondering about what the non-moral facts are. And yet they still puzzle over whether what they are doing (or what they would do in the relevant circumstances) is right. They decide to take the time to think through the issue, and that decision is driven by a final desire to do what is right *as such*.

Both philosophers and non-philosophers regularly choose to engage in moral deliberation, even while they take themselves to know all of the morally relevant non-moral facts. The best explanation of why they do this invokes a final desire to do the right thing *as such*. Because choosing to engage in moral deliberation in these cases is rational, and because such a choice is driven by a final desire to do the right thing *as such*, these are cases where having a final desire to do the right thing *as such* is rational. This shows that it is at least sometimes rational to have PMM.

## 1.2 Gathering Novel Information

Consider someone who has never felt pain before. Let's call him *Painfree*. Painfree has just come across an opportunity to torture some puppies. He thinks this would be fun. But he stops to ponder whether it would be right to do this. As he reflects, the village's moral expert swoops in and informs him that there is an experience he has never had—pain—that the puppies would experience were he to torture them. (Suppose Painfree already has very good reason to believe the village's moral expert is in fact a moral expert.) The moral expert cannot tell Painfree anything about what the experience of pain is like other than it is unlike any other experience he has ever had. But he does tell Painfree that what the experience is like bears on whether it is right for him to torture the puppies for fun. Unfortunately, right before he can tell Painfree *how* it bears on whether it is right, he gets hit by a bus.

Suppose then that the village sadist, having overheard Painfree's conversation with the moral expert, walks up to Painfree and offers him the opportunity to learn what pain feels like. After deliberating for a moment, Painfree decides to accept the sadist's offer. Why would Painfree do

this? The most plausible explanation of why he accepts the sadist's offer is that he wants to know whether it is right to torture puppies for fun, and he believes on the basis of the moral expert's testimony that learning what pain feels like will put him in a better position to gain this knowledge.

But that is not the complete explanation. Because presumably Painfree doesn't want to gain this knowledge for its own sake. He wants to gain this knowledge so that he may act on it. And he wants to act on it so that he can get something else he wants: namely, that he does the right thing, whatever that is. Ultimately, in explaining why Painfree takes the opportunity to learn how pain feels, we need to assume he has a desire to do the right thing *as such*.

This desire to do the right thing *as such* is an instance of PMM only if it's a *final* desire to do the right thing *as such*. Is it? It need not be. Painfree might want to do the right thing in order to avoid punishment, or as a way of doing what his mother wants, or to confirm that he is going to heaven. But if Painfree is virtuous, he wants it for its own sake. A virtuous Painfree has a *final* desire to do the right thing *as such*.

We can now invoke the same argument as before. Because it is rational for our virtuous Painfree to take the opportunity to experience pain for the first time, and because what motivates him to take the opportunity is a final desire to do what is right *as such*, his final desire to do what is right *as such* must also be rational. So this is a case where it is rational to have PMM. This observation constitutes further support for the claim that it is sometimes rational to have PMM.

### 1.3 Moral Deference

Now consider a case in which Painfree is unable to learn how pain feels. Perhaps he is constitutionally incapable of feeling pain. Although he cannot find out how pain feels, suppose the village's moral expert (having just jumped out of the way of the bus) tells him that the way pain feels grounds a moral obligation to refrain from torturing puppies. Having good reason to believe that the moral expert is in fact a moral expert, Painfree goes ahead and defers to the moral expert. He refrains from torturing the puppies.

What desire is Painfree acting on in deferring? He might defer to the moral expert in order to avoid punishment, or as a way of doing what his mother wants, or because he happens to just like deferring rather than making up his own mind. But if Painfree is virtuous, he defers to the moral expert because he has a desire to do what is right *as such*, and he believes that deferring to the expert will increase the probability that he will do what is right, whatever that consists in. Because Painfree is rational in deferring, and because Painfree is acting on a final desire to do

what is right *as such*, his final desire to do what is right *as such*—his PMM—must be rational. So this is another case confirming that PMM can be rational.

While Delilah's situation seemed unusual, Painfree's situation might appear rather fantastic. He is not merely uncertain which of various ways the world might be is actual. There are certain ways the world might be that he cannot even entertain. His ignorance is not merely empirical; it is conceptual.

Are we ever in the same position as Painfree? As bizarre as Painfree's case seems, real people do often face versions of it in their own lives. This occurs anytime people fail to find themselves in position to know the nature of some morally relevant non-moral fact. For example, one might think that how morally responsible an addict is for the immoral actions he takes in order to obtain drugs is at least partly a function of what it is like to be an addict. Likewise, perhaps there is something morally significant about the experience of war that one cannot even imagine until one has been in war. And perhaps one cannot understand what it is like to be the victim of various forms of oppression until one has actually been a victim.

Cases like these can be multiplied indefinitely. The important point is that we regularly defer to individuals who know certain morally relevant facts that we cannot even entertain. We defer to recovering addicts on the question of how to treat people who are in the grip of an addiction. We defer to war veterans on how we should treat prisoners of war. And we defer to people who are the targets of oppression on how they should be treated. In so doing, we are acting on a final desire to do what is right *as such*—we are acting out of PMM. And just as Painfree is rational to defer on the basis of PMM, it is rational for us to defer in these real-world cases on the basis of PMM.

I have focused on cases of moral deference where the reason for deferring to an expert is that the expert has access to morally relevant non-moral facts which one is not even in a position to entertain. But there are also cases in which one ought to defer despite the fact that the expert has no greater non-moral knowledge than oneself. For example, we often go to an impartial third party to tell us how we ought to resolve a dispute, even when the third party knows no more about the dispute than we do. Doctors often go to ethics boards to determine how they should proceed in difficult cases, even when everything the board knows about the case they learned from the doctors themselves.

What cases of this sort have in common is that they involve people rationally deferring to others who have no greater access to the morally relevant non-moral facts than they do. Because people in the cases offered above all take themselves to know all of the relevant non-moral facts, they must not be deferring to others in order to increase the probability that they perform the the act

that has the non-moral properties they care about—they are already in a position to identify which act has that feature. The best explanation of why people defer is rather that they have a final desire to do the right thing *as such*, and they believe that deferring will increase the probability that they do the right thing, whatever it is.

At this point, I take myself to have provided a considerable amount of evidence that there are circumstances in which it is rational to have PMM. Assuming there are such circumstances, any adequate view in metaethics ought to be compatible with this fact. In the rest of this paper, I argue that it is surprisingly difficult to see how any of the leading positions in metaethics could be.

## Section 2: Non-Cognitivism

According to cognitivism, moral judgments are (genuine, full-fledged) beliefs. They are in the business of describing reality, of attributing moral properties to actions, people, and states of affairs. They can be true or false in a non-deflationary, non-minimalist sense. Non-cognitivism is the denial of cognitivism.

Non-cognitivism comes in a variety of flavors. Nevertheless, all versions of non-cognitivism share one feature in virtue of which they seem to be incompatible with the rationality of PMM: they deny that we have moral concepts that represent genuine moral properties.<sup>11</sup> This feature of non-cognitivism makes it difficult for non-cognitivism to explain not only how it could be rational to have PMM, but how it could be possible to have PMM in the first place.

I have argued that PMM is best conceived of as a final desire to (e.g.) do what is right *as such*—to do the right thing *under that description*. The non-cognitivist cannot understand PMM in this way. To say that someone desires some proposition under a particular description is to say that she deploys certain concepts in order to think about that proposition. So to have a desire that one does the right thing under that description, she must deploy her concept of the property of being right (among other concepts). Importantly, this concept must *represent* the property of being right. If it didn't represent the property of being right, then it couldn't be tokened as part of a desire to do the right thing. After all, a desire is a representation of the way the world could be, a representation that plays the functional role of guiding behavior in a way that would make the world conform to that representation if one's beliefs were true. This means that a desire to do the right thing *as such* represents the world to be a certain way: the way it would be if one's act

---

<sup>11</sup> Even if the non-cognitivist can “earn the right” to speak of such concepts, a commitment to these concepts does not figure in her ground-level account of moral thought and talk. (Blackburn 1998)

instantiated the property of being right. Without a concept that represents the property of being right, we could not have a desire that represents the world to be this way, and so we could not have a desire to do the right thing *as such*.

In order to make sense of PMM within the non-cognitivist framework, then, the non-cognitivist must either reduce PMM to a desire with non-moral content, or else they must reduce it to a motivational state that lacks representational content altogether. In order to provide a solution to the puzzle of PMM, these strategies must accomplish two goals: they must identify PMM with a state that can explain why people engage in moral inquiry, and they must show that it is a state that it is rational to have. I will consider each strategy in turn.

On the first strategy, the non-cognitivist must find some final desire whose content does not involve a moral property, and this desire must be able to explain why we engage in moral inquiry. I submit that the only motivational state that satisfies this job description is a final desire to engage in moral inquiry—to deliberate, defer, and gather novel information. If the non-cognitivist attempted to identify PMM with any other desire (e.g., the desire to do what one would want to do after engaging in moral inquiry), there would be overwhelming pressure to identify the content of this desire with *doing what's right*. And this would just be to abandon non-cognitivism in favor of some form of naturalism. So the only desire the non-cognitivist can identify with PMM is a final desire to engage in moral inquiry.

If the non-cognitivist chooses to identify PMM with a final desire to engage in moral inquiry, then they will have shown that PMM is possible. But this not yet to show that it is rational. And as a matter of fact, the non-cognitivist conception of PMM does not seem rational. While it might be rational to have an instrumental desire to engage in moral inquiry, it is not rational to have a final desire to engage in moral inquiry. As entertaining as it may be, moral inquiry just isn't worth pursuing for its own sake. So although identifying PMM with the final desire to engage in moral inquiry can explain why people engage in moral inquiry, it cannot vindicate the rationality of PMM.

Now consider the non-cognitivist's second strategy for making sense of PMM. On the second strategy, the non-cognitivist must identify PMM with a motivational state that has no representational content.<sup>12</sup> In particular, they must identify it with a motivational state that has no representational content and yet can still explain why people engage in moral inquiry. What kind of state has these features? The best candidate is a brute disposition to  $\phi$  when one has a belief (construed in a manner friendly to the non-cognitivist) that  $\phi$ -ing will make it more likely that

---

<sup>12</sup> I owe this proposal to Eric Hubble, who I thank for pushing this line of response in conversation.

one does what is right.<sup>13</sup> Such a state lacks representational content, and it will explain why people engage in moral inquiry in the special case where  $\phi$ -ing is engaging in moral inquiry.

If the non-cognitivist chooses to identify PMM with such a brute disposition, they will have shown that PMM is possible. But again, that is not yet to show that it is rational. And it does not seem that a brute disposition of this sort is the kind of state that can be rational or irrational in the first place. While some brute dispositions might be rational, this does not seem to be one of them. Perhaps it is rational to have the brute disposition to believe P when you believe Q and Q is good evidence for P. But the brute disposition to engage in moral inquiry when you believe (in a manner friendly to the non-cognitivist) that doing so will lead to doing what's right is not like a disposition to respond appropriately to one's reasons. It is more like a pure associationist tendency of the mind to habitually pass from one state to another—to think of penguins after seeing purple, or to crave pickles after getting tickled.<sup>14</sup>

Suppose for the sake of argument that this disposition is rational. If this were true, then the non-cognitivist would seem to have found a solution to the puzzle of PMM. But it is a solution to the letter of the puzzle, not its spirit. For what lies behind the puzzle of PMM is the thought that engaging in pure moral inquiry is rational. So a fully general solution to the puzzle of PMM must show that pure moral inquiry is rational. It follows that the non-cognitivist will have only found a satisfactory solution to the puzzle of PMM if they can show that agents who act on the brute disposition that is PMM—agents who engage in moral inquiry—are rational. I will now argue that agents who act on this disposition are not rational.

If we grant that it is rational to have the disposition that leads people to engage in moral inquiry, what obstacle is there to concluding that it is rational to *act* on this disposition? Recall that the

---

<sup>13</sup> Or the categorical basis of such a disposition. “Brute” here is meant only to signify that the disposition is not part of the functional role of a state with representational content.

<sup>14</sup> That isn't to say that we cannot have reason to *want* to have such a brute disposition. Perhaps people who have this disposition tend to do well in business. (Suppose it's true.) So long as one has good reason to do well in business and one has good reason to believe that people who have this brute disposition tend to do well in business, one could have good reason to want to have this disposition. But having good reason to *want* to have this disposition is not having good reason to have this disposition. Someone might have good reason to *want* to believe they will succeed in some challenging endeavor, even if they have no good reason to believe this. Likewise, we might have good reason to *want* to have the brute disposition to engage in moral inquiry, but it does not follow that we have any reason to have that disposition.

disposition in question is a disposition to  $\phi$  when one believes that  $\phi$ -ing will make it more likely that one does what is right. I will argue that once we have a better understanding of what this belief is, we will see that acting on the brute disposition the non-cognitivist identifies with PMM is not rational.

I have called the state in question a “belief”. But this belief appears to have moral content. As such, the non-cognitivist cannot take it at face value. They must either reduce it to a belief with non-moral content, or else they must reduce it to a pro-attitude of some kind. On either option, the moral inquirer comes out looking irrational.

First consider the possibility that the “belief” in question is actually a pro-attitude. If it is a pro-attitude, it must be a pro-attitude toward something or other. And it seems the only thing that could be the object of this pro-attitude is the act, whatever it is, that one would perform after engaging in moral inquiry. This makes the pro-attitude into something like a final *de dicto* desire to do what one would do after engaging in moral inquiry. Unfortunately, a final *de dicto* desire to do what one would do after engaging in moral inquiry is not a desire it is rational to have. As I argue in section 6 against a specific version of analytic naturalism, it can often make sense to have an instrumental desire to do what one would do after engaging in moral inquiry (i.e., when you believe that engaging in moral inquiry increases the possibility that you will do what is *right*), but wanting to do whatever one would do after engaging in moral inquiry *for its own sake* seems downright bizarre. So if the non-cognitivist reduces the belief in question to a pro-attitude, they cannot vindicate the rationality of engaging in moral inquiry, and so fail to provide a fully general solution to the puzzle of PMM.

Now consider the possibility that the “belief” in question is a genuine belief. If it is a genuine belief, it would have to be the belief that engaging in moral inquiry increases the likelihood that one will give money to charity, or help an elderly person across the street, or . . . , where each disjunct is an action, naturalistically described, toward which one *already takes the pro-attitude required by the non-cognitivist to count as judging that action to be right*. The problem with this proposal is that it cannot explain why any rational person would bother to engage in pure moral inquiry. The proposal under consideration implies that people who engage in moral inquiry already have a settled view about which alternative is the right one. But inquiring whether P when one is already certain that P is, if not impossible, positively bizarre. It might be rational, say, for a professor who is certain utilitarianism is false to *feign* inquiry into the truth of utilitarianism for pedagogical purposes, but to genuinely wonder whether utilitarianism is true while being certain it is false cannot be rational. So again, the non-cognitivist cannot vindicate the rationality of engaging in moral inquiry, and so fails to provide a fully general solution to the puzzle of PMM.

In this section, I have considered two ways the non-cognitivist can understand PMM: as the final desire to engage in moral inquiry or the brute disposition to engage in moral inquiry. I have argued that neither proposal seems capable of accommodating the rationality of PMM—or at least the rationality of engaging in moral inquiry. If non-cognitivists are to solve the puzzle of PMM, then, they must either demonstrate that *despite appearances* one of these conceptions of PMM can vindicate the rationality of engaging in moral inquiry, or else they must offer a new conception that can.

### Section 3: Two Epistemic Constraints

In sections 4 and 5, I evaluate synthetic naturalism and non-naturalism. In the course of evaluating these views, I draw on two principles. The first principle specifies a necessary condition on rationally having a final desire for the instantiation of a property P *as such*. The second principle specifies a necessary condition on rationally having a final *preference* for the instantiation of P *as such* over the non-instantiation of P *as such*. In this section, I articulate and defend these principles.

Suppose my metaphysician friend tells me that there is some property P that he has only just discovered. He tells me nothing about P. All I am able to gather from him is that P is a property he uses the term "P" to refer to.

Suppose I then find myself overcome with a final desire that P be instantiated. And it is not merely a *de re* desire that the property be instantiated. I do not desire the instantiation of the property under the description *the instantiation of the property my friend refers to using the term "P"*. I desire it under the description *the instantiation of P*, where my concept P directly refers to the property. The desire I find myself having is a final desire for the instantiation of P *as such*. Call this desire *the Strange Desire*.

The Strange Desire is clearly irrational. But why? The Strange Desire is irrational because, even if P happens to be a perfectly desirable property, I do not know enough about P in order for me to rationally care about it for its own sake. In particular, I do not know enough about P's intrinsic nature.

So how much do I need to know about P in order to rationally have a final desire for the instantiation of P *as such*? I submit that what is needed is nothing short of knowing that property's essence.<sup>15</sup> Call this the *Epistemic Constraint on Rational Final Desire*.<sup>16</sup>

Why endorse the Epistemic Constraint on Rational Final Desire? Consider another case. Suppose that I have never before experienced the taste of vegemite. In fact, I have only just now learned that vegemite exists. Upon learning that it exists, I find myself overcome with a desire to taste vegemite *as such*.<sup>17</sup> And this isn't just any desire. It is a *final* desire. My rationale for wanting to taste vegemite is not that doing so will increase the probability that some other desire of mine will be satisfied. My rationale is that *it's vegemite*.

Is it rational for me to have the Vegemite Desire? It is not. It is not rational because I do not know enough about what tasting vegemite is in itself—I do not know enough about the intrinsic nature of the experience. All I know about the intrinsic nature of the experience is that it is a taste experience. I need to know more—I need to know something about the vegemite taste—before I can rationally desire it for its own sake.

How much more? Suppose I learn a bit more about the intrinsic nature of tasting vegemite without learning its essence—I learn that tasting vegemite necessarily involves tasting saltiness. Can I rationally have a final desire to taste vegemite *as such* once I know this much?

---

<sup>15</sup> To be more precise, I must know that property's essence, and I must know that it is that property's essence. But for simplicity I will speak only of "knowing a property's essence".

<sup>16</sup> What is the essence of a property? As I am thinking of it here, the essence of a property is what makes that property the property it is. It is the property's complete intrinsic nature. The complete intrinsic nature of a simple property is a matter of what it is in itself. The complete intrinsic nature of a complex property is a matter of what simple properties and relations it is made up of, and what those properties and relations are in themselves. Importantly, necessary features of a property may not be part of its essence. It may be a necessary feature of goodness that God loves it without God loving it being part of its complete intrinsic nature. Likewise, it may be a necessary feature of rightness that  $2+2=4$  is true, without the truth of  $2+2=4$  being part of the complete intrinsic nature of rightness. (Cf. Fine 1994.)

<sup>17</sup> Assume my concept of vegemite directly refers to vegemite. It is in virtue of deploying this concept that I manage to have a desire to taste vegemite *as such* rather than, e.g., a desire to taste what everyone actually refers to using the term "vegemite".

No. At best, I could rationally have a final desire to taste something salty *as such*. And a final desire to taste something salty *as such* is not the same as a final desire to taste vegemite *as such*.

To see why I am only in a position to rationally have a final desire to taste something salty *as such*, consider how I would go about explaining *why* I came to desire to taste vegemite after learning that it is salty. If asked why I desire to taste vegemite, I could not rationally say that it is because it is *vegemite*. The only rational answer I could give is that I desire to taste vegemite because it is *salty*. And this reveals that my desire to taste vegemite is not a final desire. It is an instrumental desire based on my final desire to taste something salty and my means-end belief that tasting vegemite is a way of tasting something salty.

While the Epistemic Constraint on Rational Final Desire is plausible, it is also reasonable to think it is too demanding. Although there is more to be said on its behalf, I will postpone discussion until section 7. (Needless to say, I think it is true enough for my purposes.) For now, let me turn to a different, less controversial principle that I can rely on should the Epistemic Constraint on Rational Final Desire turn out to be implausible. Although the arguments I give in the next section are easiest to state using the Epistemic Constraint on Rational Final Desire, versions of the arguments can also be stated using this less controversial principle. The principle can be motivated with another version of the vegemite case.

Suppose I have just learned not only that vegemite exists, but that there is another substance—marmite—that also exists. Let's say I know just as much about the intrinsic nature of marmite as I know about the intrinsic nature of vegemite. That is to say, I know that marmite tastes salty, and that's it.

Now suppose that while I desire to taste vegemite, I have no desire whatsoever to taste marmite. What should we say about this pattern of desires? We should say that is irrational. If I have a desire for the one, then I am rationally required to have a desire for the other. Moreover, if I have a desire for the one, then my desire for the other should be of exactly equal strength. After all, any rationale I might supply for desiring the one would be an equally good rationale for desiring the other.

This suggests the following principle is true: for any two properties X and Y, if you cannot distinguish the intrinsic nature of X from the intrinsic nature of Y, you cannot rationally finally desire the instantiation of X *as such* more than you finally desire the instantiation of Y *as such*. Call this the *Epistemic Constraint on Rational Final Preference*.

Having articulated and provided some initial motivation in favor of these two principles, I will now use these principles to argue that both synthetic naturalism and non-naturalism have significant trouble explaining the rationality of PMM.

## Section 4: Synthetic Naturalism

According to synthetic naturalism, moral properties are identical to natural properties, and statements identifying moral properties with natural properties are synthetic, a posteriori truths.<sup>18</sup> The most prominent version of synthetic naturalism is Cornell Realism, and the most prominent defender of Cornell Realism is Richard Boyd. I will argue here that the principles defended in the previous section show that synthetic naturalist views such as Boyd's cannot easily account for the rationality of PMM.

Boyd models his metasemantics for moral concepts on the externalist metasemantics for natural kind concepts championed by Kripke and Putnam. On this view, a concept has the referent that it does in virtue of standing in an appropriate causal relation to its referent.

A key upshot of his view is that, just as we can have the concept of "water" without knowing that the essence of water is H<sub>2</sub>O, we can possess a concept that refers to the property of rightness without knowing the essence of that property: "we can and do refer to things such that we certainly don't intend to refer to them under anything like the descriptions which in fact identify their true natures." (2003, p. 549) Moreover, we will not know the essence of the property of being right until we reach the end of moral inquiry: "The question of just which properties and mechanisms belong in the definition of [the property of being right] is an a posteriori question—often a difficult theoretical one." (1988, p. 323)

This is typically taken to be a positive feature of the view, for it enables Boyd to accommodate Moorean open questions. As Moore observed, for any natural property N, it appears to be an open question whether anything that has that property is good. This is supposed to be problematic for a naturalist view on which the identity of goodness with N is analytic. If the identity of goodness with N is settled by the meaning of "goodness" and "N", how could people who are competent with the concepts expressed by these terms question the identity of goodness and N?

Open questions are no embarrassment for Cornell Realism, for Cornell Realism predicts that competent speakers would find even true identity claims to have the appearance of being open. On this view, the identity of N with goodness is taken to be a synthetic, a posteriori truth, not a conceptual truth: "If the good is denned by a homeostatic phenomenon the details of which we still do not entirely know, then it is a paradigm case of a property whose 'essence' is given by a

---

<sup>18</sup> Like most, I do not have a good account of what makes a property a natural property. For those who lack an intuitive grip, we can think of natural properties as those properties that are properly studied by the natural sciences. (Shafer Landau 2003, p. 59)

natural rather than a stipulative definition." (1988, p. 336) Because one can refer to the property of being right (or in this case, the property of being good) without knowing its essence, someone who hasn't reached the end of inquiry can competently doubt that any given N constitutes its essence, even if it does.

The problem for Cornell Realism should by now be obvious. The feature of Cornell Realism that allows it to accommodate Moorean open questions is ultimately a bug. It is a bug because, together with the Epistemic Constraint on Rational Final Desire, it falsely implies that PMM is not rational.

The argument is simple. According to the Epistemic Constraint on Rational Final Desire, it is never rational to have a final desire for the instantiation of some property P *as such* if you do not know the essence of P. According to Cornell Realism, no one knows the essence of the property of being right. (And no one will, until they have reached the end of moral inquiry.) Therefore, if Cornell Realism is true, it seems that is never rational for people engaged in moral inquiry to have a final desire to do the right thing *as such*. Since it is rational for people engaged in moral inquiry to have a final desire to do the right thing *as such*, this poses a serious problem for Cornell Realism.

For those who are wary of the Epistemic Constraint on Rational Final Desire, I will now draw on the Epistemic Constraint on Rational Final Preference to argue for the same conclusion. The argument invoking this principle is a bit more involved, but the result is the same.

Consider the property of being right. According to Cornell Realism, we can have a final desire to do the right thing *as such*. And we can have this desire even if we don't know the essence of the property of being right. Let's suppose that the Epistemic Constraint on Rational Final Desire is false, and that it is even *rational* to have this desire. Still, the Epistemic Constraint on Rational Final Preference threatens to undermine the rationality of engaging in pure moral inquiry. For whatever the property of being right is, there will be a very similar property being right\* which is just like being right, except that its essence does not include  $\phi$ -ing. Whatever  $\phi$ -ing is, we can imagine a case where someone does not know whether  $\phi$ -ing is right, and wants to do what is right. Now we can ask this person: why not desire to do what's right\* instead? After all, doing what's right\* and doing what's right appear to you to be indistinguishable. You may know that  $\phi$ -ing lies within the essence of one and not the other, but you don't know whether it lies within the essence of doing what's right\* or whether it lies within the essence of doing what's right. So why, then, prefer to do the right thing rather than the right\* thing? As far as final preferences go, what rationality requires in this case is indifference. This is both the intuitive result, and the result we get by applying the Epistemic Constraint on Rational Final Preference.

Actual moral inquirers prefer doing what's right to doing what's right\*. What the foregoing shows is that they would not be rational in having this preference if they did not know how doing what's right differed from doing what's right\*. Since Cornell Realism implies that people do not know how doing what's right differs from doing what's right\*, Cornell Realism implies that people are not rational in having this preference. Since people are rational in having this preference, this poses a serious challenge for Cornell Realism.

The problem for synthetic naturalism is that it seems to deny that ordinary moral inquirers can know the essence of a moral property without knowing the complete truth about what lies in its extension. This problem is particularly acute for Cornell Realism given how closely it ties the essence and extension of moral properties. If synthetic naturalists want to solve the PMM, they must tease the essence of moral properties apart from their extension. Moreover, they must show that, despite being synthetic, truths giving the essences of moral properties are generally available to ordinary moral inquirers before they engage in significant philosophical investigation. It remains to be seen whether there is a version of synthetic naturalism that can satisfy these desiderata while retaining the benefits traditionally associated with synthetic naturalism.

## Section 5: Non-Naturalism

Non-naturalism is a form of cognitivism, but it denies that moral properties are identical with natural properties. Instead, it takes moral properties to be a species of *non-natural* properties. Is non-naturalism compatible with the rationality of PMM?

Some have thought that non-natural properties cannot be worth caring about. Here is a characteristic remark from Frank Jackson:

Are we supposed to take seriously someone who says, 'I see that this action will kill many and save no-one, but that is not enough to justify my not doing it; what really matters is that the action has an extra property that only ethical terms are suited to pick out'? In short, the extra properties would [be] ethical 'idlers'. (1998, p. 127)

Some non-naturalists reply to Jackson's challenge by claiming that they never took non-natural moral properties to be worth caring about in the first place.<sup>19</sup> Still, we can find non-naturalists who care a great deal about non-natural moral properties. For example, in *On What Matters* Derek Parfit repeatedly claims that if there are no non-natural moral properties, "Sidgwick, Ross, and I, and others would have wasted much of our lives." (2011 p. 12; p. 304; p. 367) We should

---

<sup>19</sup> See, e.g., Fitzpatrick (2014, p. 564).

take seriously, then, the idea that there are non-natural moral properties whose essence makes them worth caring about for their own sake.

Unfortunately, other commitments that non-naturalists like Parfit have make it difficult to take this idea seriously. Here, for example, is one of Parfit's central quietist commitments:

- (V) There are some claims that are irreducibly normative in the reason-involving sense, and are in the strongest sense true. But these truths have no ontological implications. For such claims to be true, these reason-involving properties need not exist either as natural properties in the spatio-temporal world, or in some non-spatio-temporal part of reality. (Parfit 2011, p. 486)

So by Parfit's own lights, *there is no sense* in which the properties he cares about so deeply exist. In light of the fact that they don't exist, it is hard to make sense of the idea that they might have an essence that makes them worth caring about. If this is right, then quietism seems unable to accommodate the rationality of PMM because it fails to identify moral properties with properties that are worth caring about for their own sake.<sup>20</sup>

There are other non-naturalists, however, who are willing to countenance the idea that non-natural properties exist and have an essence that makes them worth caring about for their own sake. For example, David Enoch, a self-styled "robust realist", thinks there is a real sense in which non-natural moral properties exist and matter (2011). Can robust realism solve the PMM?

If robust realism is to avoid the problems faced by non-naturalism and quietism, then non-natural moral properties must satisfy two desiderata. First, everyday people must know their essence. Second, the essence of these non-natural moral properties must allow them to be worth caring about for their own sake.

Consider the second desideratum first. Given how little non-naturalists have been able to say about the essence of non-natural moral properties, there is still the residual Jacksonian worry that whatever these properties are like, they just can't matter. At this point, the best bet for the non-naturalist is to give up the project of describing the essence of being right in independent terms and simply claim that the essence of being right, ineffable as it is, makes it worth caring about for its own sake. Just as someone pressed to explain why they desire pleasure for its own sake can do no more than point to the nature of the experience itself, someone pressed to explain why

---

<sup>20</sup> Parfit isn't alone in his quietism. For example, Scanlon seems to be committed to the idea that we cannot identify the property picked out by the basic normative concept of a reason (2014, p. 44).

they desire to do what's right for its own sake may be able to do no more than point to the nature of the the property of being right.<sup>21</sup> "To know it is to love it," they might say.

However plausible this move is, let us now consider the first desideratum. Can the robust realist make sense of how ordinary moral inquirers could come to know the essence of a non-natural moral property?<sup>22</sup> In order to answer this question, the non-naturalist must show that it is possible to stand in a special relation of acquaintance with that property, a relation that gives one knowledge of its essence.<sup>23</sup> Moreover, they must make it plausible that ordinary moral inquirers actually stand in this relation to non-natural moral properties.

What would such an essence-revealing relation look like? The most vivid example of such a relation comes from visual experience. Articulating a view he attributes to Bertrand Russell, Mark Johnston writes that “one naturally does take and should take one's visual experience as of, e.g. a canary yellow surface, as completely revealing the intrinsic nature of canary yellow, so that canary yellow is counted as having just those intrinsic and essential features which are evident in an experience as of canary yellow.” (1992, p. 223) Likewise, to use an earlier example, it is plausible to think that the essence of the taste of vegemite is completely revealed to someone who experiences the taste of vegemite.

The case of canary yellow provides a model for how the essence of a property can be revealed through experience. But non-naturalists generally resist the idea that moral properties are the

---

<sup>21</sup> Cf. Moore (1903, p. 7).

<sup>22</sup> If the Epistemic Constraint on Rational Final Desire is false but the Epistemic Constraint on Rational Final Preference is true, then all the robust realist needs to show is that we can distinguish the intrinsic nature of doing right from the intrinsic nature of doing right\*. It should be clear how to amend the arguments below to show that the robust realist cannot do this.

<sup>23</sup> The kind of acquaintance that can deliver knowledge of an entity's essence more closely resembles Russellian acquaintance than the kind of acquaintance familiar from debates about direct reference and singular thought. I can be acquainted with an entity in this latter sense without knowing its essence. The kind of acquaintance I have in mind corresponds to that which Dasgupta (2015) has in mind: “Let us say that a subject S is *acquainted* with *x* iff the nature of *x* is directly presented or revealed to S (this is just a label, not an analysis).” (p. 464)

kinds of properties we can be immediately acquainted with in conscious experience.<sup>24</sup> So what other option is there?

Perhaps standing in a special causal relation to an entity can give one knowledge of its essence, even if that causal relation doesn't give rise to any conscious experience of the object's intrinsic nature. Unfortunately, non-naturalists generally hold that moral properties are causally inert.<sup>25</sup> So this proposal is also a non-starter for standard versions of non-naturalism.

It would be too quick to conclude at this point that we don't know the essence of non-natural moral properties just because they are causally inert and because we are not immediately acquainted with them. After all, it is plausible to think we know the essence of certain logical relations and mathematical entities, but it is not plausible to think that we are causally connected to or immediately acquainted with them (Dasgupta 2015, p. 464-465; Chalmers 2012, p. 404). So how does one come to know the essence of a logical or mathematical entity? A natural proposal is to hold that we know these items by having logical and mathematical concepts that play an appropriate role in our cognitive economy (Chalmers 2012, p. 466). Can a similar claim be made to substantiate the view that we know the essence of non-natural moral properties? Whether such a story could be told depends on whether an appropriate conceptual role can be found for our moral concepts. I will consider two proposals.

On the first proposal, the appropriate conceptual role consists in being able to apply the concept in exactly those cases in which it is correctly applied. Put another way, this proposal holds that one might know the essence of a moral property by knowing the complete truth about what lies

---

<sup>24</sup> For exceptions, see Oddie (2005) and Atiq (manuscript). Atiq explores such a view without explicitly endorsing it. Oddie argues that we are acquainted with value in the experience of desire. It is unclear whether he thinks that acquaintance with value gives us knowledge of value's essence. See also Johnston (2001) who endorses the existence of "the affective disclosure of value" (p. 183).

<sup>25</sup> For exceptions, see Shafer-Landau (2012), Wedgwood (2007), Cuneo (2006) and Oddie (2005). Although these views allow for the causal efficacy of moral entities, more work needs to be done to show that this causal efficacy allows us to know the essence of any moral entity. For discussion of whether we might be unable to know the essence of even causally efficacious entities, see Langton (1998), Lewis (2009), Locke (2009), and Dasgupta (2015).

in its extension across all possible worlds.<sup>26</sup> Unfortunately, this proposal is a non-starter, for the cases where PMM appears most clearly to be rational—cases of people engaging in moral inquiry—are precisely cases where people lack knowledge of what lies in the extension of the property. So the non-naturalist who accepts this proposal cannot deliver the verdict that PMM is rational.

On the second proposal, the appropriate conceptual role has less to do with our disposition to apply the concept and more to do with how we behave once we've applied the concept. Indeed, Ralph Wedgwood and David Enoch have recently offered such accounts in an attempt to explain how we can come to think about non-natural moral properties in the first place (Wedgwood 2007; Enoch 2011). Can such an account also explain how we could have knowledge of a moral property's essence?

No, and it is easy to see why without digging too far into the details of these views. On these views, a particular mental representation refers to a particular non-natural moral property in virtue of playing a certain “downstream” conceptual role. For example, on Wedgwood's view, I count as competent with the concept of *being what I ought to do* just in case believing that I ought to  $\phi$  commits me (in the relevant sense) to  $\phi$ -ing. This is a very easy condition to satisfy. Indeed, it does not require that I have any true beliefs about what I ought to do. I could believe the only thing I ever ought to do with my time is torture puppies for fun and yet still count as competent with the concept of *being what I ought to do*, so long as in believing that I ought to torture puppies I commit myself (in the relevant sense) to torturing puppies. (Cf. Schroeter & Schroeter 2009, pp. 4-9; Gibbard 2003, pp. 28-29, pp. 149-150)

This case might not pose any threat to the idea that Wedgwood's view is correct as an account of how our thoughts manage to latch onto the non-natural moral property of *being what I ought to do*. But it poses an insuperable problem for the idea that this same mechanism gives competent users of the concept knowledge of its referent's essence. This is because it is highly implausible to think that I could believe that the only thing I ever ought to do is torture puppies while at the

---

<sup>26</sup> This proposal goes naturally with a “pleonastic” conception of properties (Schiffer 2003), but it is not inconceivable to imagine a robust realist endorsing it. For example, Shafer-Landau (2003) endorses a version of non-naturalism on which every instance of a non-natural moral property is exhaustively constituted by natural properties. It seems that on Shafer-Landau's view, knowing the complete story about which combinations of natural properties suffice to constitute a non-natural moral property would be enough to know the essence of that moral property.

same time knowing the essence of *being what I ought to do*. Although it is plausible to think that knowing the essence of *being what I ought to do* is compatible with not being able to tell what has that property in *hard* cases, it is not at all plausible to think that someone could know the essence of that property while simultaneously believing that all they ought to do is torture puppies.

In order to solve the puzzle of PMM, the robust realist needs to show how it is possible for ordinary people know the essence of non-natural moral properties. I have considered several proposals for how this could be and found them all inadequate. In order to solve the puzzle of PMM, then, the non-naturalist must either show that my dissatisfaction with these proposals is misplaced, or else revise their view so as to allow ordinary moral inquirers to know their essence.

## Section 6: Analytic Naturalism

According to analytic naturalism, moral properties are identical to natural properties, and the relevant identity statements are analytic truths. Different versions of the view identify different moral properties with different natural properties, but all versions of the view agree that the correct identities are analytic truths.

The leading variant of analytic naturalism is a version of moral functionalism, defended by Frank Jackson and Philip Pettit (Jackson 1998; Jackson and Pettit 1996; Jackson and Pettit 1995). According to analytic moral functionalism, the property of being right (e.g.) is the property of having the property that plays the rightness role in mature folk morality, and this is an analytic truth.

This requires a bit of unpacking. Folk morality is the "the network of moral opinions, intuitions, principles and concepts whose mastery is part and parcel of having a sense of what is right and wrong, and of being able to engage in meaningful debate about what ought to be done." (Jackson 1998, p. 130) *Mature* folk morality is "where folk morality will end up after it has been exposed to debate and critical reflection (or would end up, should we keep at it consistently and not become extinct too soon)." (p. 133) Mature folk morality consists in a set of claims about various moral phenomena. From these claims, we can extract a role associated with each moral property. We can extract a rightness-role, a wrongness-role, a goodness-role, a badness-role, and so on for each of the moral properties. To take an example, the rightness-role might look something like this: being shared by all the acts that mature folk morality judges to be right, being related to other moral properties in the way mature folk morality takes them to be, and being such as to influence behavior in the way mature folk morality assumes it would. Suppose that there is some property that makes these claims (or at least a sufficient number of them) true. Call this the

*realizer property*. According to analytic moral functionalism, the property of being right is analytically equivalent to the property of having the realizer property—the property of having the property that plays the rightness-role in mature folk morality.<sup>27</sup>

Unlike Cornell Realism and non-naturalism, moral functionalism allows that everyday people who are competent with moral concepts know the essence of the properties those concepts refer to. They may not be able to articulate this knowledge in terms of necessary and sufficient conditions, but they nevertheless have it as tacit knowledge. So moral functionalism succeeds where Cornell Realism and non-naturalism fail.

But moral functionalism faces a problem of its own: although it allows ordinary inquirers to know the essences of moral properties, it takes the essences of these properties to be such that that we have no reason to care about them for their own sake. For example, it identifies the property of being right with the property of having the property that plays the rightness-role in mature folk morality. Whether an act instantiates the property of having the property that plays the rightness-role in mature folk morality does not matter for its own sake, if it matters at all. If someone tells me that  $\phi$ -ing has the property of having the property that plays the rightness-role in mature folk morality, that provides me with no reason whatsoever to  $\phi$ . The fact that a certain act has a property that a certain group of (future or hypothetical) people *treat* as right does not itself give us any reason to perform that act.

Thus, moral functionalism fails to accommodate the rationality of PMM because it identifies moral properties with properties that are not worth caring about for their own sake. Since moral properties are worth caring about for their own sake, this poses a serious problem for moral functionalism.

Is there any way of saving moral functionalism? Perhaps the moral functionalist would deny that when we desire to do what is right, we are thereby desiring that we do what has the property of having the property that plays the rightness role in mature folk morality... *under that description*. Instead, perhaps we are desiring it under a description that provides a substantive

---

<sup>27</sup> In their initial statement of the view, Jackson and Pettit remain neutral on whether *rightness* should be identified with the property of having the realizer property or the realizer property itself. If we understand rightness to be the property of being right, then the latter interpretation makes moral functionalism into a version of synthetic naturalism, a version that fails for the same reason that Cornell Realism fails: it implies that ordinary moral inquirers don't know the essence of *being right*. In his most recent statement of the view, Jackson explicitly identifies the property of being right with the property of having the realizer property (or in other terms, "the role property"). (2016, p. 207)

account of what playing the rightness role amounts to. What might that be? Here we can take up something Jackson says about what it is to *believe* that something is right. According to Jackson, "to believe that something is right is to believe in part that it is what we would in ideal circumstances desire," where being what we would in ideal circumstances desire is part of what it is to have a property that plays the rightness role in mature folk morality (1998, p. 159). This suggests that to desire that we do the right thing is to desire in part that *we do what we would in ideal circumstances desire*,<sup>28</sup> where those ideal circumstances include being fully informed, coherent, etc.

The problem remains: it is not rational to have a final desire that one's act has the property of being what we would in ideal circumstances desire. It may make sense to care *instrumentally* about the fact that an act would be what we would in ideal circumstances desire—after all, this fact strongly suggests that the act has a property worth caring about for its own sake. But it makes no sense to care *for its own sake* about the fact that  $\phi$ -ing would be what we would in ideal circumstances desire. This is especially clear when we recall that an ideal version of ourselves might fail to want what we know to be right: "Even if one is initially benevolent, complete awareness of the suffering of the mass of sentient beings would be horrifically depressing, and hardness of heart rather than valuing their release might well be the causal upshot." (Johnston 1989, p. 152)

It seems that analytic moral functionalism cannot accommodate the rationality of PMM. If moral functionalists are to solve the puzzle of PMM, they must show that playing the rightness role in mature folk morality amounts to something that is worth caring about for their own sake. In general, analytic naturalists who want to solve the puzzle of PMM must identify moral properties with properties that are worth caring about for their own sake. Whether there is a plausible version of analytic naturalism that can do this remains to be seen.

---

<sup>28</sup> A move like this makes moral functionalism into a type of ideal observer or ideal advisor theory (e.g., Firth 1952; Smith 1994).

## Section 7: An Objection

The Epistemic Constraint on Rational Final Desire has played a large role in my arguments. Unfortunately, it appears to face counterexamples.<sup>29</sup> For example, if it is rational to have a final desire for any particular thing *as such*, it is rational to have a final desire for knowledge *as such*. But the lesson of the literature on the Gettier problem is that we do not know the essence of knowledge. So if the Epistemic Constraint on Rational Final Desire is correct, final desires for knowledge *as such* are not rational. But they are rational. So the constraint must be false.<sup>30</sup>

In response, I have two options: deny that final desires for knowledge are ultimately rational, or else provide reason to think we actually do know the essence of knowledge. I will take the second.

First, why think that we don't know the essence of knowledge? Sure, the attempt to articulate a set of necessary and sufficient conditions for knowledge has been a spectacular failure. But this doesn't show that we don't know the essence of knowledge. In fact, the Gettier literature suggests quite the opposite. For nearly every case that has come up in the literature on the Gettier problem, people have a strong intuition about whether knowledge obtains. This shows we have an ability to determine whether knowledge obtains across various counterfactual scenarios. And this in turn strongly suggests that we have at least tacit knowledge of knowledge's essence. And this, I suggest, is all that is required to make it rational to want knowledge for its own sake.

So my response to the objection is that knowing the essence of a property doesn't require being able to state necessary and sufficient conditions for its instantiation. The required knowledge can be tacit.<sup>31</sup>

---

<sup>29</sup> Versions of these alleged counterexamples can also be cooked up for the Epistemic Constraint on Rational Final Preference. It should be obvious how the replies I offer below could be adapted to defend the Epistemic Constraint on Rational Final Preference from the alleged counterexamples.

<sup>30</sup> This objection was raised to me in conversations with Chris Howard and Gideon Rosen.

<sup>31</sup> Importantly, as argued in section 5, embracing this account of what it takes to know a property's essence will not show that ordinary moral inquirers know the essences of moral properties.

That being said, there are other cases where it seems rational to have a final desire for the instantiation of P *as such* even though one does not have even tacit knowledge of P's essence. For example, most people want Mom to be happy *as such*. (Suppose "Mom" directly refers to one's own mother.) Even supposing people know the essence of happiness, no one knows the essence of their mother. And yet people have a final desire that Mom is happy *as such*, and they seem to be completely rational in having this desire. This appears to be a counterexample to the Epistemic Constraint on Rational Final Desire.

In reply, all I wish to note is that even if it is rational to have a final desire that Mom is happy *as such*, the rationale for this desire must involve facts about Mom's extrinsic nature. After all, if asked why one wants Mom to be happy, one's reply will be something like "Mom gave birth to me" or "Mom raised me" or "Mom and I are pretty close to each other". These replies all invoke facts about Mom's extrinsic properties. Giving birth to someone is not part of anyone's intrinsic nature. Nor is raising someone or being close to them. This suggests a natural revision to the Epistemic Constraint; instead of holding that it is rational to have a final desire for the instantiation of some property P *as such* only if one knows the essence of that property, we should say that it is rational to have a final desire for the instantiation of some property P only if one knows the essence of that property and *facts about that property's extrinsic nature do not constitute any part of one's rationale for having that desire*.

This might seem like a significant concession. But it does nothing to undermine the use to which I've put the Epistemic Constraint. This is because the exceptions to the Epistemic Constraint that we've identified are all cases where facts about the desired property's extrinsic nature constitute some part of one's rationale for having that desire, and facts of this kind play no part in any virtuous person's rationale for having PMM. If asked why they desire to do the right thing, neither Delilah or Painfree would cite anything about the extrinsic nature of rightness—they wouldn't say "Because my mother wants me to do the right thing." They would say "Because it's right." So even if there are exceptions to Epistemic Constraint on Rational Final Desire, PMM isn't one of them.

## Section 8: Conclusion

I have argued that standard versions of non-cognitivism, analytic naturalism, synthetic naturalism, and non-naturalism have trouble vindicating the rationality of PMM. Although I believe the prospects for non-cognitivism are bleak, I am optimistic that some plausible form of cognitivism will be developed that can solve the puzzle of PMM. While it is not clear what that view will look like, we know in advance that it will have to have two features that the views considered here lack. First, it will have to allow that everyday people know the essence of moral

properties despite not knowing the complete story about what lies in their extension. Second, the essences it assigns to moral properties must make those properties worth caring about for their own sake.

As a matter of fact, it is easy to find views that have both of these features. For example, consider a simple version of analytic naturalism according to which the property of being right is identical with the property of maximizing pleasure. Assuming people know the essence of maximizing pleasure, this view explains how people can know the essence of being right. It also identifies the property of being right with a property that is worth caring about for its own sake. As a result, it implies that PMM is rational and so solves the problem of PMM. Even so, this view is widely thought to be eminently implausible, and being able to solve the puzzle of PMM won't change that. This is because solving the puzzle of PMM is just one of many desiderata a metaethical account ought to satisfy, and if a view fails to satisfy too many other desiderata, the fact that it solves the puzzle of PMM will do little to redeem it.

That being said, some desiderata are more important than others. How important is solving the puzzle of PMM? I submit that it is very important. Indeed, I take the assumption that moral inquiry is worthwhile—and thus that PMM is rational—to play a foundational role in fixing the subject matter of metaethics. I contend that the role it plays is so important that an “account of morality” that fails to capture this assumption is ultimately an account of something else altogether. So while many desiderata could be jettisoned if it turned out that satisfying them was especially costly, solving the puzzle of PMM could not be. Solving the puzzle of PMM is non-negotiable.

## References

- Arpaly, N. (2002). Moral Worth. *Journal of Philosophy*, 99 (5), 223–245.
- Atiq, E. H. (unpublished manuscript). Intuitionism, Humeanism, and the Metaphysics of Desire.
- Blackburn, S. (1998). *Ruling Passions*. Oxford: Oxford University Press.
- Boyd, R. N. (2003). Finite Beings, Finite Goods: The Semantics, Metaphysics and Ethics of Naturalist Consequentialism, Part I. *Philosophy and Phenomenological Research*, 66(3), 505-553.
- Boyd, R. N. (1988). How to Be a Moral Realist. In G. Sayre-McCord (Ed.), *Essays on Moral Realism*, 181-228.
- Chalmers, D. (2012). *Constructing the World*. Oxford: Oxford University Press.
- Cuneo, T. (2006). Moral Facts as Configuring Causes. *Pacific Philosophical Quarterly*, 87 (2), 141–162.
- Dasgupta, S. (2015). Inexpressible Ignorance. *Philosophical Review*, 124 (4), 441–480.
- Dreier, J. (2000). Dispositions and Fetishes: Externalist Models of Moral Motivation. *Philosophy and Phenomenological Research*, 61 (3), 619–638. <http://doi.org/10.2307/2653615>
- Enoch, D. (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
- Fine, K. (1994). Essence and Modality. *Philosophical Perspectives*, 8, 1–16.
- Firth, R. (1952). Ethical Absolutism and the Ideal Observer. *Philosophy and Phenomenological Research*, 12(3), 317-345.
- Fitzpatrick, W. (2014). Skepticism about Naturalizing Normativity: In Defense of Ethical Nonnaturalism. *Res Philosophica*, 91 (4), 559-588.
- Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, 23 (6), 121-123.
- Gibbard, A. (2003). *Thinking How to Live*. Harvard University Press.
- Jackson, F. (2000). *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- Jackson, F. (2016). In Defense of Reductionism in Ethics. In P. Singer, *Does Anything Really Matter?* (pp. 195-211). Oxford: Oxford University Press.

- Jackson, F. & Pettit, P. (1995). Moral Functionalism and Moral Motivation. *Philosophical Quarterly*, 45 (178), 20-40.
- Jackson, F., & Pettit, P. (1996). Moral Functionalism, Supervenience, and Reductionism. *Philosophical Quarterly*, 46 (182), 82-86.
- Johnston, M. (1989). Dispositional Theories of Value, III. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 63, 139-174.
- Johnston, M. (1992). How to Speak of the Colors. *Philosophical Studies*, 68 (3), 221–263.
- Johnston, M. (2001). The Authority of Affect. *Philosophy and Phenomenological Research*, 63(1), 181–214.
- Langton, R. (1998). *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford University Press.
- Lewis, D. (2009). Ramseyan Humility. In D. Braddon-Mitchell & R. Nola (Eds.), *Conceptual Analysis and Philosophical Naturalism* (pp. 203–222). Cambridge, MA: MIT Press.
- Locke, D. (2009). A Partial Defense of Ramseyan Humility. In D. Braddon-Mitchell & R. Nola (Eds.), *Conceptual Analysis and Philosophical Naturalism* (pp. 223-241). Cambridge, MA: MIT Press.
- Lord, E. (2015). Acting for the Right Reasons, Abilities, and Obligation. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics*, vol. 10. Oxford University Press.
- McGrath, S. (2009). The Puzzle of Pure Moral Deference. *Philosophical Perspectives*, 23 (1), 321–344.
- Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.
- Oddie, G. (2005). *Value, Reality, and Desire*. Oxford: Oxford University Press.
- Parfit, D. (2011). *On What Matters: Volume Two*. Oxford: Oxford University Press.
- Scanlon, T. M. (2014). *Being Realistic about Reasons*. Oxford: Oxford University Press.
- Schiffer, S. (2003). *The Things We Mean*. Oxford: Oxford University Press.
- Schroeter, L., & Schroeter, F. (2009). A Third Way in Metaethics. *Noûs*, 43(1), 1–30. <http://doi.org/10.1111/j.1468-0068.2008.01693.x>
- Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford: Oxford University Press.

Shafer-Landau, R. (2012). Evolutionary Debunking, Moral Realism, and Moral Knowledge. *Journal of Ethics and Social Philosophy*, 7 (1).

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

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

Wedgwood, R. (2007). *The Nature of Normativity*. Oxford: Oxford University Press.