

6

The Puzzle of Pure Moral Motivation

Adam Lerner

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.¹

Some people who have a desire to do what's right *as such* desire to do what's 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's the right thing. They have a *final* desire to do what's right *as such*. This desire is an instance of what I will call *pure moral motivation* (PMM).² Other instances of PMM include the desire to avoid doing what's wrong *as such*, the desire to promote what's good *as such*, and the desire to prevent what's bad *as such*.

Contemporary views in metaethics have been built to satisfy 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 many others. In this chapter, I introduce a new desideratum: vindicating the rationality of PMM. I argue that PMM is often rational, 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 the standard views in metaethics.³

In Section 6.1, I argue that PMM is at least sometimes rational. In Section 6.2, I explain why non-cognitivism has trouble accommodating

¹ The terminology originates with Smith (1994, pp. 74–6).

² McGrath (2009) provides the inspiration for both the terminology and the chapter title.

³ While I focus on *moral* motivation in this chapter, I believe the puzzle arises for every kind of pure normative motivation.

the rationality of PMM. In Section 6.3, I introduce two principles regarding the conditions under which final desires and preferences are rational. In Sections 6.4 and 6.5, I draw on these principles to argue that synthetic naturalism and non-naturalism both have trouble accommodating the rationality of PMM. In Section 6.6, I explain why analytic naturalism has trouble accommodating the rationality of PMM. Section 6.7 concludes.

6.1. THE RATIONALITY OF PMM

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 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 seek the moral truth. They might get paid to do it. They might want to predict what virtuous people will do in certain circumstances so they may exploit their virtuous character. They might just be curious.

But this is not why most people seek the moral truth. Most people seek the moral truth because they want to act in accordance with the moral truth. They want to do the right thing *as such*.⁴ And they want to do it for its own sake. They have a *final* desire to do the right thing *as such*.⁵ They exhibit PMM.

In this section, I argue that it can be rational to have PMM.⁶ I offer two types of cases in which it's rational to choose to engage in moral inquiry and in which the best explanation of why ordinary people do this is that they

⁴ 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's conceived to be.

⁵ 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) (cf. Harman, 2000; McDaniel and Bradley, 2008).

⁶ For purposes of this chapter, a motivation is rational if and only if the person who has the motivation possesses at least *some* good reason of the right kind to have that motivation and it's their possession of this reason that sustains that motivation. For someone to possess a reason to have some motivation, it isn't enough that there be some reason for them to have it: that reason must be within their ken (cf. Lord, 2015).

have PMM. I argue that because in both types of cases the choice to engage in moral inquiry is rational, the PMM that lies behind this choice must itself be rational.

6.1.1. Pure Moral Inquiry with Full Information

Delilah owns a factory farm. She knows everything there is to know about what goes on inside a factory farm. She also knows everything there is to know about how factory farms impact the environment, the economy, and people's gustatory experiences. But Delilah has never thought about whether it might be wrong to run a factory farm. 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 time and effort into trying to figure out whether running a factory farm is right?

Suppose Delilah chooses to deliberate. She seems perfectly rational in making this choice. Why would Delilah make this choice? Clearly, she has a desire to know what's morally right in this case. But *presumably* she doesn't want this bit of moral knowledge for its own sake. She wants it because she wants to *do* what's right.⁷

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

So when Delilah chooses to deliberate, she must be acting on a desire to do what's right *as such*.⁸ Now suppose she doesn't 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's bound for heaven. She wants it for its own sake. Her desire to do what's right is a final desire. This shows that in choosing to deliberate, Delilah is acting on PMM.

Is it rational for Delilah to have this PMM? Yes. Why? Because Delilah acts rationally in choosing to deliberate. And if an agent rationally chooses to

⁷ Or at least, she wants to avoid doing what's wrong.

⁸ Alternatively, she may be acting on a second-order final (*de dicto*) desire to do whatever final desires constitute (*de re*) desires to do what's right (Dreier, 2000, p. 632). Or she may be acting on a desire to do what's *just* or *benevolent*, as long as justice and benevolence each are irreducibly moral or have an irreducibly moral component. Since these alternative desires essentially have moral content, they are still instances of PMM. So what I go on to say about the *de dicto* desire to do what's right applies just as much to these desires.

φ , then the desires she acts on in making that choice are rational.⁹ Thus, we have a case in which PMM is rational.¹⁰

A Humean might object that this merely shows Delilah isn't irrational in acting on her PMM. They might hold that PMM is a deep-seated final desire that cannot itself be subject to rational assessment. I disagree. It seems it would be rational for Delilah to act on PMM even if she knew that doing so would likely defeat an even more deep-seated desire of hers (e.g., getting rich). What this shows is that when we say Delilah's PMM is rational, we're not just saying it's immune to rational criticism. We're saying that it's positively rational—that she has some good reason to have that desire.

Delilah's case may sound strange. Does any actual person ever take themselves to know all of the relevant non-moral facts and still engage in moral inquiry by deliberating or seeking out pure moral testimony?

Indeed, pure moral inquiry of this kind is perfectly familiar. When people find themselves puzzling over whether it's okay to get an abortion, whether they're giving enough to charity, or whether they should stop eating meat, they aren't *always* just wondering about how the world is in non-moral respects. They often *take* themselves to know all of the relevant non-moral facts. Likewise, people often go to an impartial third party to tell them how to resolve a dispute, even when the third party knows no more about the dispute than they do. In such cases, people rationally choose to engage in moral inquiry, and the only plausible explanation of why they do so is that they have PMM. So Delilah's case is not all that unusual. PMM is common and often rational.

6.1.2. Pure Moral Inquiry with Conceptual Limitations

Painfree has never felt pain before. He has just come across an opportunity to torture some puppies, and 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—whose nature bears on whether it's right to torture puppies. (Suppose Painfree already has excellent reason to believe the village's moral expert is a moral expert.) The moral expert cannot tell Painfree anything about the intrinsic nature of the experience. But the

⁹ While cases where it's rational to desire that one act (or be disposed to act) irrationally are familiar from the literature on rational irrationality, I know of no cases in which someone acts rationally (i.e., responds adequately to their reasons) but in so doing acts on an irrational desire.

¹⁰ Although my argument does not depend on it, I also believe Delilah is praiseworthy (cf. Smith, 1994; Arpaly, 2002; Weatherson, 2014).

expert does tell Painfree that what the experience is like bears on whether it's right to torture puppies. Unfortunately, just before the expert can tell Painfree *how* it bears on whether it's right, the expert gets hit by a bus.

Suppose then that the village sadist—having overheard Painfree's conversation with the expert—walks up and offers Painfree the opportunity to learn what pain feels like, free of charge. After deliberating, Painfree decides to accept the sadist's offer. Why would Painfree do this? He isn't trying to figure out whether "pain" refers to something he antecedently cares about—he doesn't know anything about pain, so he hasn't any opinion about it. In accepting the sadist's offer, then, he must be trying to figure out whether "pain" refers to something that it's *wrong to cause*. Why does he want to know this? He must want to know this because he wants to do what's right. Now suppose that Painfree does not think that doing what's right *as such* would promote the satisfaction of any of his other desires. It follows that his desire must be a final desire. Because a final desire to do what's right is an instance of PMM, this shows that Painfree must have PMM.

It also shows that Painfree's PMM is rational. Why? Because Painfree acts rationally in accepting the sadist's offer. And because Painfree acts on PMM in accepting the sadist's offer, his PMM must be rational. So we have another case where PMM is rational.

While Delilah's situation seems unusual, Painfree's situation might appear fantastical. He is not merely uncertain which of various non-normative ways the world might be is actual. There are certain non-normative ways the world might be that he cannot even conceive. His ignorance is not empirical; it's conceptual.

Are we ever in Painfree's position? Yes. This occurs whenever we know there is some morally relevant non-moral fact whose nature we don't know. For example, one might think that how morally responsible an addict is for the wrongs they commit in order to obtain drugs is partly a function of what it's like to be an addict, and that one cannot know this without having been an addict. Likewise, perhaps there are morally relevant facts about living through war or being oppressed that one cannot even imagine until one has been to war or been oppressed.

In cases like these, we regularly defer to individuals who know morally relevant facts that we cannot even conceive. 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 oppressed on how they should be treated. In so doing, we are acting on a final desire to do what's right *as such*—we are acting out of PMM. And just as Painfree is rational to defer on the basis of PMM, so are we.

At this point, I take myself to have provided compelling evidence that there are circumstances in which it's rational to have PMM. I assume in

what follows that any adequate metaethics ought to account for this fact—and I argue that it's surprisingly difficult to see how any of the standard views in metaethics could. Those who remain unconvinced by the arguments in this section can interpret what follows as an argument for a conditional claim: if PMM is rational, then it's hard to see how any of the standard views in metaethics could accommodate that fact.

6.2. NON-COGNITIVISM

According to cognitivism, moral judgments are (genuine, full-fledged) beliefs. They're 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 many flavors. But all pure versions of non-cognitivism share one feature in virtue of which they are incompatible with the rationality of PMM: they deny that we have moral concepts that represent genuine moral properties.¹¹ This shared assumption makes it impossible for non-cognitivism to explain how it could be rational to have PMM since it implies that it's not possible to have PMM in the first place.¹² The reason why is that PMM is a final desire to (e.g.) do what's right *as such*—to do the right thing *under that description*—and it is thereby partly constituted by a moral concept that represents a genuine moral property. Since non-cognitivism denies the existence of such concepts, it must deny the existence of PMM.

This shows non-cognitivism cannot solve the puzzle of PMM, narrowly construed. But it does not show that it cannot solve the puzzle of PMM, broadly construed. The puzzle of PMM, broadly construed, is to explain how it can be rational to engage in (pure) moral inquiry. I argued that we must assume people act on PMM because this provides the best explanation of why people engage in moral inquiry. The non-cognitivist denies the existence of PMM. In its place, they must find some other mental state—let us call it *non-cognitivist PMM* or *NCPMM*—that it's rational to have and

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

¹² For purposes of this chapter, I assimilate hybrid versions of non-cognitivism (e.g., Ridge, 2014; Toppinen, 2013; Schroeder, 2013) to the naturalistic views I go on to discuss. While they may be able to account for the possibility of PMM, it's unclear whether they can account for the rationality of PMM. Thanks to Mike Ridge for discussion.

that can play the same role as PMM in explaining why people engage in moral inquiry. Only by identifying such a state can they solve the puzzle of PMM, broadly construed.

In order to evaluate whether NCPMM satisfies these two desiderata, we need a better idea of what it could be. The non-cognitivist has two options. They can either identify NCPMM with a desire that has exclusively non-moral content, or else they can identify it with a motivational state that lacks representational content altogether. I will consider each strategy in turn. I will argue that while both can explain why people engage in moral inquiry, neither can explain why people are *rational* in doing so.

On the first strategy, the non-cognitivist must find some final desire whose content doesn't 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. If the non-cognitivist attempted to identify NCPMM with any other desire (e.g., the desire to do what one would want to do after engaging in moral inquiry or to do what one's true self wants to do), 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 NCPMM is a final desire to engage in moral inquiry.

If the non-cognitivist chooses to identify NCPMM with a final desire to engage in moral inquiry, then they will have shown that NCPMM can explain why people engage in moral inquiry. But in order to solve the puzzle of PMM, broadly construed, they must show that it's rational to have this state. And in fact, it doesn't seem rational to have this state. While it might be rational to have an *instrumental* desire to engage in moral inquiry—as a way of doing something intellectually stimulating, or as a way of increasing the probability of doing what's right—it isn't rational to have a *final* desire to engage in moral inquiry.

Now consider the second strategy. On this strategy, the non-cognitivist identifies NCPMM with a motivational state that has no representational content whatsoever and yet can still explain why people engage in moral inquiry.¹³ Such a state would have to be a brute disposition to φ when one has the belief that φ -ing will increase the probability that one does what's right.¹⁴

If the non-cognitivist chooses to identify PMM with such a brute disposition, they will have shown that NCPMM can explain why people

¹³ Thanks to Eric Hubble for offering this line of response in conversation.

¹⁴ Or the categorical basis of such a disposition.

engage in moral inquiry. But again, in order to solve the puzzle of PMM they must show that it's rational to have this state. And it doesn't seem rational to have this state, because it doesn't seem that this brute disposition is the kind of state that it can be rational *or* irrational to have in the first place. While some brute dispositions might be rational to have, this doesn't seem to be one of them. Perhaps it's 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 that doing so will lead you to do what's right is not like a disposition to respond appropriately to one's reasons. It's more like a pure association, a mere tendency to pass from one mental state to another. While it might be rational to try to cultivate in oneself brute dispositions of this sort, the dispositions themselves are not rational or irrational. There is no right kind of reason to have states like these.

But suppose for the sake of argument that it's rational to have this brute disposition. If this were true, then the non-cognitivist would have found a state that it's rational to have and that can explain why people engage in moral inquiry. But that is not yet to solve the puzzle of PMM, broadly construed. In order to do that, the non-cognitivist must show that not only having this brute disposition but *acting* on it—i.e., engaging in moral inquiry—is rational. I will now argue that they cannot do this.

If we grant that it's rational to have the disposition in question, what obstacle is there to concluding that it's rational to *act* on this disposition? Recall that the disposition in question is a disposition to φ when one has the *belief* that φ -ing will increase the probability that one does what's right. I will argue that once we have a better understanding of what this belief is according to non-cognitivism, we will see that acting on this brute disposition 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 reduce it either to a genuine belief with non-moral content, or else 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 a pro-attitude. As 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 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 and other desires like this are not desires it's rational to have. As I argue in Section 6.6 against a 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, but it's bizarre to want to do whatever one would do after engaging in moral inquiry *for its own sake*. 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 solve the puzzle of PMM, broadly construed.

Now consider the possibility that the belief in question is a genuine belief with non-moral content. If so, 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 moral inquiry. Why? Because it implies that people who engage in moral inquiry already have a settled view about which of their options they ought to take. But this makes people who engage in moral inquiry look irrational, for it means they already have an answer to the question they pursue in moral inquiry. Genuinely inquiring whether P when one is already certain that P is, if not impossible, irrational. So again, the non-cognitivist cannot vindicate the rationality of engaging in moral inquiry, and so fails to solve the puzzle of PMM, broadly construed.

6.3. AN EPISTEMIC CONSTRAINT

In Sections 6.4 and 6.5, I evaluate synthetic naturalism and non-naturalism. In evaluating these views, I draw on the following principle:

The Epistemic Constraint on Rational Final Desire. For any property P, if you do not know P's essence, it is not rational to finally desire that P be instantiated *as such*.¹⁵

Why endorse this Epistemic Constraint? 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

¹⁵ What is the essence of a property? For these purposes, 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's 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 essence (cf. Fine, 1994).

desire to taste Vegemite *as such*.¹⁶ And this isn't just any desire. It's a *final* desire. My rationale for wanting to taste Vegemite isn't that doing so will increase the probability that some other desire of mine will be satisfied. My rationale is that *it's tasting Vegemite*.

Is it rational for me to have the Vegemite Desire? It isn't. Why? Because I don't know enough about what it's like to taste Vegemite—I don't know enough about the intrinsic nature of the experience. All I know about it is that it's 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 Vegemite's taste without learning its essence—I learn that Vegemite's taste is, among other things, salty. Can I rationally have a final desire to taste Vegemite *as such* once I know this much?

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

To see why I'm 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 desire to taste Vegemite. If asked why I desire to taste Vegemite, I could not rationally say that it's because it's *Vegemite*. The only rational answer I could give is that I desire to taste Vegemite because it tastes *salty*. And this reveals that my desire to taste Vegemite is not a final desire. It's 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. This same line of reasoning can be repeated for any version of the case in which my knowledge of how Vegemite tastes falls short of knowing its essence. Furthermore, it can be repeated for any case in which I desire something whose essence I don't know. There's nothing special here about the nature of subjective experience. Whenever I lack knowledge of an object's essence, my desire for that object can only be instrumental if it is to be rational.¹⁷

It may seem that the Epistemic Constraint on Rational Final Desire makes rational final desire hard to come by. For example, many think that if it's rational to have a final desire for anything *as such*, it's rational to have a

¹⁶ Assume my concept of Vegemite directly refers to Vegemite. It's 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."

¹⁷ Experiencing the taste of Vegemite can allow one to know the essence of that experience only if property dualism is true. Even if property dualism is false, the Epistemic Constraint is compatible with the existence of a rational final desire in the vicinity: the final desire to have whatever experience I know under such-and-such a mode of presentation (i.e., the phenomenal character of tasting Vegemite).

final desire for knowledge *as such*. But the lesson of the literature responding to Gettier's seminal (1963) article is that we don't 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. Since they are rational, the Epistemic Constraint on Rational Final Desire faces an apparent counterexample.

Whether this objection succeeds depends on whether we know the essence of what we finally desire. Take the case of knowledge. The failure of the Gettier literature doesn't show that we don't know the essence of knowledge. People have strong intuitions about whether knowledge obtains in Gettier cases. If this ability to determine whether knowledge obtains extends to *all* possible cases, then we seem to have at least tacit knowledge of knowledge's essence. And this may be all that's required to make final desires for knowledge rational.

But even if our ability to determine whether knowledge obtains does *not* extend to all possible cases, we may still know the essence of knowledge in some other way, so long as its essence does not consist in what falls within its extension. If we don't—if the essence of knowledge consists in what falls within its extension and we don't know what falls within its extension—then the Epistemic Constraint seems to get the right result: it isn't rational to finally desire knowledge. If it were, it would be rational for someone to finally desire to have knowledge in some case, to not know whether they have knowledge in that case, and to know all of the other facts about that case. But clearly someone who knew all the facts that could ground their having knowledge (whether their belief was true, reliably formed, etc.) and yet did not know whether they have knowledge could not rationally have a final desire that they have knowledge. This person would irrationally desire they know-not-what.

Even so, there are other cases where it clearly seems rational to have a final desire for *P as such* even though one clearly does not know *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 Mom. And yet people have a final desire that Mom is happy *as such*, and they seem to be rational in having this desire.

This shows there is a class of exceptions to the Epistemic Constraint. What unifies this class? Notice that the rationale for the Mom Desire must involve facts about Mom's extrinsic nature. If asked why one wants Mom to be happy, one's reply must be something like "Mom gave birth to me" or "Mom raised me" or "Mom and I are pretty close." These replies all invoke facts about Mom's extrinsic properties. This suggests that you can be rational in having a final desire for an object when you don't know

the essence of that object *so long as you know the essence of that object's extrinsic nature and the object's extrinsic nature makes the object worth caring about for its own sake.*

Having noticed that there is this class of exceptions to the Epistemic Constraint, we can safely ignore it. In what follows, I argue that synthetic naturalism and non-naturalism cannot vindicate the rationality of PMM because the Epistemic Constraint is true and these views imply that we don't know the essence of any moral property. Even though the Epistemic Constraint has exceptions, it isn't plausible to resist this argument by saying that PMM constitutes such an exception. This is because facts about the extrinsic nature of moral properties play no part in any virtuous person's rationale for having PMM. If asked why they desire to do the right thing, neither Delilah nor 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 the Epistemic Constraint, PMM isn't one of them.

Although the arguments in Sections 6.4 and 6.5 are stated most succinctly using this principle, I will now offer a second, less controversial principle that can do the same work. This is the Epistemic Constraint on Rational Final Preference: for any two properties X and Y, if you cannot distinguish the intrinsic nature of X from the intrinsic nature of Y, you cannot remain rational while finally desiring the instantiation of X *as such* more than you finally desire the instantiation of Y *as such*. So if I have just learned not only that Vegemite exists, but that there is another substance—Marmite—that also exists, and all I know about the intrinsic nature of either is that they're salty, then I cannot rationally desire to taste one more than the other. This result is eminently plausible, and its plausibility does not turn on anything specific about taste sensations. So the Epistemic Constraint on Rational Final Preference poses an equally severe problem for the standard views in metaethics: the problem of explaining why it's rational to want to do what's right more than what's not right.

6.4. 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. The most prominent version of synthetic naturalism is Cornell Realism, and the most prominent defender of Cornell Realism is Richard Boyd. I will argue that our Epistemic Constraints 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 H₂O, we can possess a concept that refers to the property of being right 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” (Boyd, 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.” (Boyd, 1988, p. 197)

Boyd takes this to be a positive feature of the view, for it enables it 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 allegedly 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 a synthetic, a posteriori truth, not a conceptual truth: “If the good is defined 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 natural rather than a stipulative definition” (Boyd, 1988, p. 210) 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’s 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’s never rational to have a final desire for the instantiation of some property P *as such* if you don’t 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 reach the end of moral inquiry.)

Therefore, if Cornell Realism is true, it's never rational for ordinary people engaged in moral inquiry to have a final desire to do the right thing *as such*. Since it's rational for ordinary people engaged in moral inquiry to have this desire, this poses a serious problem for Cornell Realism.

I have just argued against Cornell Realism relying on the Epistemic Constraint on Final Desire. But the Epistemic Constraint on Rational Final Preference poses an equally severe problem for Cornell Realism: that of explaining why it's rational to want more strongly to do what's right *as such* than to do what's not right *as such*. To see why, consider that 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 doesn't include φ -ing. Suppose we have an agent who is in a "hard case" and is thus unsure whether φ -ing is right. Suppose this person wants to do what's right *as such* for its own sake. Now we can ask this person: why not desire to do what's right* for its own sake instead? After all, doing what's right* and doing what's right appear to you to be indistinguishable. You may know that φ -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 the essence of doing what's right*. So why, then, finally prefer to do the right thing rather than the right* thing? If the Epistemic Constraint on Rational Final Preference is true, what rationality requires in this case is indifference. But since it *does* seem rational to finally prefer doing what's right over what's right*, this poses a serious problem for Cornell Realism.

6.5. 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?

If non-naturalism is to allow for the rationality of PMM, then non-naturalism must avoid the problem faced by synthetic naturalism; it must allow everyday people to know the essence of moral properties. But it must also avoid a familiar objection to non-naturalism: that non-natural moral properties are not worth caring about for their own sake. These two desiderata must be satisfied if non-naturalism is to solve the puzzle of PMM.

Consider the second desideratum first. Some have thought that non-natural properties cannot be worth caring about for their own sake. 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 for their own sake in the first place.¹⁸ This reply, however, simply concedes that non-naturalism cannot solve the puzzle of PMM. Fortunately, we can find non-naturalists who believe non-natural moral properties are worth caring about. 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, pp. 12, 304, 367) But why think non-natural moral properties have an essence that makes them worth caring about?

At this point, the best bet for the non-naturalist is to give up on describing the essence of being right and to just 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 they desire to do what’s right for its own sake may be able to do no more than point to the nature of being right.¹⁹ “To know it is to love it,” they might say.

However plausible this move is, it will not allow the non-naturalist to solve the PMM. And that is because non-naturalism fails the first desideratum considered above: it cannot explain how ordinary moral inquirers could know the essence of a non-natural moral property. In order to show that this desideratum is satisfied, the non-naturalist must show that it’s possible to stand in a special relation of acquaintance with that property, a relation that delivers knowledge of its essence.²⁰ Moreover, they must show that ordinary moral inquirers actually stand in this relation to non-natural moral properties. I will argue that standard versions of non-naturalism cannot show this.

What would an essence-revealing relation of acquaintance 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

¹⁸ See, e.g., FitzPatrick (2014, p. 564).

¹⁹ Cf. Moore (1903, p. 7).

²⁰ The kind of acquaintance I have in mind corresponds to what Dasgupta (2015, p. 464) 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).”

essential features which are evident in an experience as of canary yellow” (1992, p. 223). Likewise, it’s plausible to think that the essence of Vegemite’s taste is completely revealed to someone who tastes Vegemite.

The case of canary yellow provides a model for how the essence of a property can be revealed through experience. But standard versions of non-naturalism reject the idea that moral properties are the kinds of properties we can be acquainted with in conscious experience.²¹ 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 nature. Unfortunately, standard versions of non-naturalism hold that moral properties are causally inert.²² 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. After all, it’s plausible to think we know the essence of certain logical relations and mathematical entities, but it isn’t plausible to think we’re causally connected to or immediately acquainted with them (Dasgupta, 2015, pp. 464–5; Chalmers, 2012, p. 404). So how does one come to know the essence of a logical or mathematical entity? A natural proposal is that we have this knowledge by having logical and mathematical concepts that play an appropriate role in our cognitive economy (Chalmers, 2012, p. 466). Can a similar proposal show 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’s correctly applied. This proposal holds that one might know a moral property’s essence by knowing the complete truth about what lies in its extension across all possible worlds. Unfortunately, this proposal is a non-starter, for the cases where PMM appears most clearly rational—cases of people engaging in moral inquiry—are precisely cases where people lack knowledge of what lies in the property’s extension. So non-naturalists who accept this proposal cannot deliver the verdict that PMM is rational.

²¹ For possible exceptions, see Oddie (2005), Atiq (unpublished manuscript), and Johnston (2001).

²² For exceptions, see Shafer-Landau (2012), Wedgwood (2007), Cuneo (2006), and Oddie (2005). Although these views allow for causal efficacy, more work needs to be done to show that this causal efficacy allows us to know the essence of any moral entity (cf. Langton, 1998; Lewis, 2009; Locke, 2009; and Dasgupta, 2015).

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 it. Indeed, Ralph Wedgwood and David Enoch have 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 accounts also explain how we could know a moral property's essence?

No. On these views, a mental representation refers to a 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 φ commits me (in the relevant sense) to φ -ing. This is a very easy condition to satisfy. Indeed, it doesn't require that I have any true beliefs about what I ought to do. I could believe the only thing I ever ought to do 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 and Schroeter, 2009, pp. 4–9; Gibbard, 2003, pp. 28–29, 149–50).

Even if 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*, the present case shows that this account cannot also show that we know that property's essence. This is because it's implausible to think that I could believe that the only thing I ever ought to do is torture puppies while at the same time knowing the essence of *being what I ought to do*. Although it's plausible to think that knowing the essence of *being what I ought to do* is compatible with not knowing what has that property in *hard* cases, it isn't at all plausible to think that someone could know that property's essence while simultaneously believing that all they ought to do is torture puppies. Knowing a moral property's essence can't come *that* far apart from knowledge of what lies in its extension.

I have argued that non-naturalism cannot vindicate the rationality of PMM because it cannot show how ordinary moral inquirers could know any moral property's essence. This argument goes through only if the Epistemic Constraint on Rational Final Desire is true. But the same conclusion can be reached using the Epistemic Constraint on Rational Final Preference. The problem is that non-naturalism has a hard time explaining how we could learn how the essence of being right differed from, say, the essence of being right*. The typical routes by which we would come to know this difference are closed off: we aren't acquainted with them in experience, and we don't causally interact with them. And it's hard to see how having a concept that plays the role specified by Wedgwood could tell us anything about the essence of being right—or at least anything that wasn't also shared by being

right*. Lastly, we might know that the action we are inquiring about falls into the extension of the one but not the extension of the other, but we don't know which—if we did, we wouldn't have to deliberate. So even if we reject the Epistemic Constraint on Rational Final Desire, the Epistemic Constraint on Rational Final Preference creates an equally serious problem for standard versions of non-naturalism.

6.6. 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 agree that the correct identities are analytic truths.

The leading variant of analytic naturalism is moral functionalism, defended by Frank Jackson and Philip Pettit (Jackson, 1998; Jackson and Pettit, 1995). According to 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.²³

Unlike Cornell Realism and non-naturalism, moral functionalism allows that everyday people who are competent with moral concepts know the essence of the properties these concepts refer to. They may not be able to articulate a set of necessary and sufficient conditions that captures this essence, but they nevertheless have tacit knowledge of it. 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. This is because it identifies the property of being right with the property of having the property that plays the rightness role in mature folk morality.

According to Jackson, folk morality is “the network of moral opinions, intuitions, principles and concepts whose mastery is part and parcel of having a sense of what's right and wrong, and of being able to engage in

²³ In their initial statement of the view, Jackson and Pettit remain neutral on whether *rightness* should be identified with the property of having the property that plays the rightness role (i.e., the property of having the *realizer property*) or with the property that plays the rightness role (i.e., 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*.

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)” (Jackson, 1998, p. 133). So, roughly, the property of being right is the property we would *treat* as right if we took our ordinary sense of right and wrong and cleaned it up by thinking and arguing for a long time. Unfortunately, whether an act instantiates this property doesn’t seem to matter for its own sake. Why should we care for its own sake about doing what we would *treat* as right under such conditions?

A satisfactory answer to this question would have to come by way of a richer description of what it would be to treat something as right under such conditions, which Jackson supplies. According to Jackson, “to believe that something is right is to believe in part that it is what we would in ideal circumstances desire,” and 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 desire if we were fully informed, coherent, etc. as such.*²⁴ And doesn’t it make sense to care for its own sake about whether you’re doing what you would desire under such ideal conditions?

No. 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 distinct property that’s worth caring about for its own sake. But it makes little sense to care *for its own sake* about the fact that φ -ing would be what we would in ideal circumstances desire. At least, it doesn’t make sense to care as strongly about this fact as it does to care about whether we do what’s right. Caring so strongly about whether our act satisfies the desires we would have in ideal conditions *for its own sake* seems fetishistic. So if moral functionalism identifies our strong desire to do what’s right with a strong desire to do what we would in ideal circumstances desire, it can’t explain why it would be rational for us to engage in moral inquiry as frequently as we rationally engage in moral inquiry.

6.7. CONCLUSION

I have argued that the standard views in metaethics all have trouble vindicating the rationality of PMM. Although the prospects for non-cognitivism

²⁴ A move like this makes moral functionalism into a type of ideal observer or ideal advisor theory (e.g., Firth, 1952; Smith, 1994).

and synthetic naturalism are bleak, some plausible form of analytic naturalism or non-naturalism may yet be developed that can solve the puzzle of PMM. While it isn't clear what that view will look like in all its details, we know in advance that it will have two features that no standard view has both of. First, unlike standard versions of synthetic naturalism and non-naturalism, it will have to allow that everyday moral inquirers know the essence of moral properties despite not knowing whether the acts they inquire about lie in their extension. Second, unlike standard versions of analytic naturalism, the essences it assigns to moral properties must make those properties worth caring about for their own sake.

It remains an open question whether there is an adequate view that can satisfy both of these constraints. But we have reason to be optimistic. Since each constraint is satisfied by at least one of the standard views, we know that each constraint can be satisfied in principle. And we have no reason to doubt they can both be satisfied by a single view.

Suppose it turns out that these constraints can't both be satisfied by a single view. We then have a choice. We can either conclude these are not genuine constraints, or we can take the fact that no view satisfies them as a new argument for error theory about morality. I myself would be inclined toward error theory. But that is an argument for another day.²⁵

References

- Arpaly, N. (2002). Moral Worth. *Journal of Philosophy*, 99 (5), 223–45.
- Atiq, E. H. (unpublished manuscript). Intuitionism, Humeanism, and the Metaphysics of Desire.
- Blackburn, S. (1998). *Ruling Passions*. Oxford: Oxford University Press.
- Boyd, R. N. (1988). How to Be a Moral Realist. In G. Sayre-McCord (Ed.), *Essays on Moral Realism* (pp. 181–228). Ithaca, NY: Cornell 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), 505–53.
- Chalmers, D. (2012). *Constructing the World*. Oxford: Oxford University Press.
- Cuneo, T. (2006). Moral Facts as Configuring Causes. *Pacific Philosophical Quarterly*, 87 (2), 141–62.

²⁵ Thanks to audiences at Princeton University, the University of Michigan, the Central European University, and the 2016 Chapel Hill Metaethics Workshop. Special thanks to David Heering, Eric Hubble, Mark Johnston, Sarah McGrath, Joseph Moore, Mike Ridge, Gideon Rosen, Thomas Schmidt, Michael Smith, Nat Tabris, and two anonymous reviewers for extensive discussion and written feedback.

The Puzzle of Pure Moral Motivation

143

- Dasgupta, S. (2015). Inexpressible Ignorance. *Philosophical Review*, 124 (4), 441–80.
- Dreier, J. (2000). Dispositions and Fetishes: Externalist Models of Moral Motivation. *Philosophy and Phenomenological Research*, 61 (3), 619–38.
- 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), 317–45.
- FitzPatrick, W. (2014). Skepticism about Naturalizing Normativity: In Defense of Ethical Non-Naturalism. *Res Philosophica*, 91 (4), 559–88.
- Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, 23 (6), 121–3.
- Gibbard, A. (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Harman, G. (2000). Desired Desires. In *Explaining Value* (pp. 117–36). Oxford: Oxford University Press.
- Jackson, F. (1998). *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- Jackson, F. and Pettit, P. (1995). Moral Functionalism and Moral Motivation. *Philosophical Quarterly*, 45 (178), 20–40.
- Johnston, M. (1992). How to Speak of the Colors. *Philosophical Studies*, 68 (3), 221–63.
- 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: Oxford University Press.
- Lewis, D. (2009). Ramseyan Humility. In D. Braddon-Mitchell and R. Nola (Eds.), *Conceptual Analysis and Philosophical Naturalism* (pp. 203–22). Cambridge, MA: MIT Press.
- Locke, D. (2009). A Partial Defense of Ramseyan Humility. In D. Braddon-Mitchell and R. Nola (Eds.), *Conceptual Analysis and Philosophical Naturalism* (pp. 223–41). 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 (pp. 26–52). Oxford: Oxford University Press.
- McDaniel, K., and Bradley, B. 2008. Desires. *Mind*, 117 (466), 267–302.
- McGrath, S. (2009). The Puzzle of Pure Moral Deference. *Philosophical Perspectives*, 23 (1), 321–44.
- 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.
- Ridge, M. (2014). *Impassioned Belief*. Oxford: Oxford University Press.
- Schroeder, M. (2013). Tempered Expressivism. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics*, vol. 8 (pp. 283–313). Oxford: Oxford University Press.
- Schroeter, L., and Schroeter, F. (2009). A Third Way in Metaethics. *Noûs*, 43 (1), 1–30.
- Shafer-Landau, R. (2012). Evolutionary Debunking, Moral Realism, and Moral Knowledge. *Journal of Ethics and Social Philosophy*, 7 (1), 1–37.

- Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.
- Toppinen, T. (2013). Believing in Expressivism. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics*, vol. 8 (pp. 252–82). Oxford: Oxford University Press.
- Weatherson, B. (2014). Running Risks Morally. *Philosophical Studies*, 167 (1), 141–63.
- Wedgwood, R. (2007). *The Nature of Normativity*. Oxford: Oxford University Press.