Non-Naturalistic Moral Explanation

Sam Baron (University of Western Australia)

Kristie Miller (University of Sydney)

Michael Rubin (University of Western Australia)

Jonathan Tallant (University of Nottingham)

1. Introduction

It has seemed, to many, that there is an important connection between the ways in which some theoretical posits explain our observations, and our reasons for being ontologically committed to those posits. One way to spell out this connection is in terms of what has become known as the explanatory criterion of ontological commitment. This is, roughly, the view that we ought to posit only those entities that are indispensable to our best explanations. The motivation for a criterion such as this is clear: it aims to rule out commitment to 'ontologically dubious' entities—entities such as undetectable fairies at the bottom of one's garden. The explanatory criterion is sometimes framed as a fairly strong thesis: that we ought (epistemically) to posit *all and only* those entities that are indispensable to the best available explanations of our observations.[1]

It is easy to see, however, that even a relatively weak version of the thesis, such as the thesis that we ought (epistemically) to posit *only* those entities that are indispensable to the best available explanations of our observations, will cause problems for a range of philosophical views. In particular, various non-naturalistic philosophical theories

---

[1] This explanatory thesis is thus a refinement of the Quinean indispensability thesis: we ought (epistemically) to posit *all and only* those entities that are indispensable to our best scientific theories (Colyvan (2001).

appear, at least on the face of it, to be unable to meet even this weak version of the explanatory criterion. For it typically seems that particular naturalistic facts are sufficient to do the explanatory work: the non-naturalistic facts are dispensable to any explanation. If that's right, however, then non-naturalistic accounts of morality, reason, prudence, and normativity more generally, are all at risk, since there is simply no explanatory work for the non-natural facts to do. Since the weak criterion is sufficient to raise these problems, henceforth it is this criterion that we have in mind when we talk of *the* explanatory criterion of ontological commitment.

This paper focuses on a particular kind of non-naturalism: moral non-naturalism. Our primary aim is to argue that the moral non-naturalist places herself in an invidious position if she simply accepts that the non-natural moral facts that she posits are not explanatory. This has, hitherto, been the route that moral non-naturalists have taken. They have attempted to make their position more palatable by pointing out that there is reason to be suspicious of the explanatory criterion of ontological commitment. That is because other perfectly respectable views fall foul of that criterion, most notably: mathematical realism. Since we don't want to rule out mathematical realism, we should jettison the explanatory criterion of ontological commitment. Against this manoeuvre, we argue that many contemporary mathematical realists accept the explanatory criterion and provide an account of how mathematical objects are indeed indispensable to our best explanations. Thus, the moral non-naturalist will be left in an awkward dialectical position if she accepts that non-natural moral properties play no such explanatory role.

This provides the impetus for the secondary aim of the paper, which is to offer the non-naturalist an alternative strategy for dealing with moral explanations. The strategy is to retain the explanatory criterion of ontological commitment and maintain that non-

natural moral properties are, in fact, explanatory. The explanations they provide, however, are non-causal explanations. We will explain how the moral non-naturalist can make use of explanations of this kind, by giving an analysis in counterfactual terms. In this way, the non-naturalist appears to recover their partner in crime, the mathematical realist, who can also understand the explanatory contribution that mathematics makes in terms of non-causal explanations underpinned by counterfactuals.

As we shall see, the strategy that we offer the non-naturalist may require her to endorse counterpossibles—counterfactuals with impossible antecedents—as the basis for understanding the explanatory contribution made by non-natural moral properties, depending on her antecedent views about the nature of such properties. While counterpossibles are controversial, we will argue that moral naturalists are committed to their non-vacuous truth just as much as non-naturalists, and so the controversy surrounding counterpossibles cannot be used against the non-naturalist in the debate over the nature of moral properties.

2. Moral Realism and Explanation

We understand moral realism to be the view that there are moral facts, and that these facts are stance-independent: the standards that determine the moral facts are not made true by the beliefs, attitudes, conventions, or agreements of actual or hypothetical appraisers.[2] In what follows, we suppose that *naturalistic* moral realism entails that moral properties are nothing over and above natural properties: that is, moral properties

---

[2] This characterization follows Shafer-Landau (2003: 15). See also Brink (1989: Ch. 2).

are reducible to, or identical with, or fully constituted by, or fully grounded in natural properties. By contrast, we suppose that *non-naturalistic* moral realism is the view that moral properties *are* something over and above natural properties: the moral properties are not reducible to, or identical with, the natural properties (though some may be reducible to other non-natural properties). The non-naturalistic moral realist typically holds that, despite the independence of the moral, the moral properties nonetheless supervene on the natural properties. The question arises as to what counts as a natural property and what counts as a non-natural property. We have no wish to adjudicate this vexed issue here. So, in what follows we will simply suppose that the natural properties are those with causal powers. The non-natural properties are those that lack causal powers.[3]

Sturgeon invokes a counterfactual test for explanatory relevance in defence of realism. According to this test, roughly, a fact *x* is explanatorily relevant to a distinct fact *y* if it is true that, if *x* had not occurred, *y* would not have occurred (Sturgeon 1985: 65). Sturgeon argues that the putative moral facts pass the counterfactual test. It is often thought (Thomson 1996) that if moral facts explain anything at all, the potential explananda are limited to the moral beliefs, actions, and attitudes or responses of agents.[4] Thus, Sturgeon offers the following example of moral explanation: Hitler's depravity—a moral property—explains why he ordered the genocide. This case is then analysed counterfactually: if Hitler hadn't been morally depraved, he would not have ordered the genocide—for only a depraved person would do such a thing; no morally decent person would deliberately initiate an ethnic genocide.

---

[3] Our characterization is in line with that of Miller (2013: 10) and Sturgeon (2003: 538).

[4] Thomson (1996: 76).

On the face of it, one might think that the non-naturalist can avail herself of the response just offered. That is because non-naturalists almost universally agree with naturalists that moral properties supervene upon natural properties, even if they disagree about exactly which natural properties these are. Since the obtaining of supervenience relations is sufficient to ground the truth of various counterfactuals, and since the naturalist and non-naturalist broadly agree about what supervenes on what, they will assent to the truth of the same kinds of counterfactuals. So, if the obtaining of those counterfactuals is sufficient to show that the moral facts are explanatorily indispensable, then the non-naturalist is as well placed to offer moral explanations as is the naturalist.

We noted earlier, however, that non-naturalists do not offer such explanations. Instead, they deny that explanatory indispensability is necessary for the justification of ontological commitment to an entity. Why? They do so because they, alongside naturalistic moral realists, accept that the obtaining of the relevant sorts of counterfactuals is evidence for, but is not constitutive of, explanatory relevance. That is because both parties accept that the reason the counterfactuals are evidence of explanatory relevance is because they are also evidence of causation. It is *causation* that is held to be explanatory, and it is causal explanations that naturalistic moral realists seek. Yet it is often supposed, as we have, that it is a mark of non-natural properties that they are causally inert.[5] What is lurking in the background, here, is the assumption that explanation is essentially causal, and so if non-natural properties are causally inert, then they are, in turn, non-explanatory.

---

[5] A view along these lines is expressed by Sturgeon (1985, 2003: 543n24), Copp (2003: 186), Scanlon (1998: 58). Railton (1989: 154) also makes a connection between natural properties and empirical (which seems to mean 'causal') explanation.

Non-naturalistic moral realists suppose, however, that they have independent reason to reject the explanatory criterion of ontological commitment, arguing that such a criterion is a poor one, since it will lead us astray in other cases, in particular with respect to mathematical realism (e.g. Scanlon (2014) and Parfit (2011)). In what follows, we will take a closer look at this partners-in-crime defence of moral non-naturalism, and argue that, due to recent developments in the philosophy of mathematics, it is no-longer a viable response.

3. Partners in Crime

According to mathematical realists, some mathematical statements are objectively and mind-independently true and are made true by the existence of abstract objects, namely numbers, sets, functions, and the like. These objects, by virtue of being abstract, cannot enter into causal networks, and are traditionally thought to not play a role in causal explanations. Non-natural moral realism would seem to be in no worse position than this. Moreover, mathematical realism is historically an important and prominent position in the philosophy of mathematics. Indeed, moral realists of all stripes have long looked towards mathematical realists as fellow travellers. Scanlon, for instance, appeals to such a partners-in-crime defence when he argues that Harman′s (1977) explanatory requirement ″does not apply to the normative domain, or to mathematics″ (Scanlon 2014: 26). [6] Mathematical realism has a commitment to seemingly spooky

---

[6] See also Putnam (2004), and Parfit (2011). For an insightful discussion of the relationship between the moral and mathematical cases in the context of epistemology, see Clarke-Doane (2014). The view we outline in this paper is somewhat different from Clarke-Doane′s 'epistemic' argument, which focuses on

abstract objects outside the causal realm, and has no convincing epistemology (Benacerraf 1983; Field 1989). Despite this, mathematical realism seems to have good credentials. If the non-naturalistic moral realist is, as it seems, backed into a corner, they at least have a friend in there with them.

There was a time when this partners-in-crime defence was plausible; but more recently, some mathematical realists (e.g., Baker 2009; Colyvan 2002, Colyvan 2010) have argued that mathematics does deliver explanations. There have been several cases advanced where it appears that mathematics is playing a crucial role in the explanation of physical phenomena. For example, it has been argued that mathematics contributes to explanations of such diverse phenomena as the Kirkwood gaps in the asteroid belt (Colyvan 2010), the hexagonal shape of honey comb cells (Lyon and Colyvan 2008), the Lorenz contraction of bodies with large relativistic velocities (Colyvan 2002), the periodic behaviour of North American cicadas (Baker 2005), the failure of attempts to square the circle (Colyvan 2007), the search patterns of open-sea aquatic predators (Baron 2014), and the instability of particular high-energy galaxies (Lyon and Colyvan 2008).

---

the epistemology of mathematics and morality. He argues that, contrary to what some have argued, it's not the case that mathematical knowledge is epistemically more tractable (via mathematical proof) than is moral knowledge. By contrast, we do not focus on the epistemology of either enterprise, but, instead, on whether both (or neither) can provide explanations for other phenomena. Notice that these are not the same, since in theory *X* could provide an explanation for *Y*, even if one does not have a good account of how one comes to know *X*. Clarke-Doane does also consider indispensability arguments in favour of moral and mathematical realism. See footnote (8) for more on this.

An example here will help. There are two subspecies of North American cicadas with prime-numbered life cycles. That is, these cicadas appear from the ground en mass once every 13 and 17 years respectively. It turns out that there is considerable evolutionary advantage in having such a life cycle. For instance, if you are trying to avoid a predator with a periodical life cycle, you need to minimise the number of years you overlap with that predator. It can be proved that having a prime life cycle does this (Lehman-Ziebarth et al. 2005; Baker 2005).[7] The idea here is that the mathematical properties of prime numbers essentially enter into the explanation of why the cicadas have their prime-numbered life cycles.

We will return to this case a bit later on. For now, it is enough to note that if mathematics does play a genuine explanatory role within science, then the moral non-naturalist stands alone in her rejection of the explanatory criterion of ontological commitment. In order for the moral non-naturalist to preserve their partners-in-crime defence, they must now argue that the most recent wave of mathematical realism is in error; despite appearances to the contrary, mathematics does not play an explanatory role in science.[8]

---

[7] See Baker (2005) for details of the proof. It is intuitively easy to see how it goes, though. If a sub-species′ life cycle leaves it vulnerable to predation once every 13 years, say, only predators with a similar and coordinated 13-year life cycle, annual predators, and biannual predators, again suitably coordinated (i.e. starting on an odd-numbered year) will overlap with it on a regular basis.

[8] In this paper we focus on the explanatory criterion for ontological commitment, and the worry that non-natural moral properties fall foul of this commitment. There is a related, but distinct, worry in the vicinity, which is that we are not justified in believing in such properties because we could not gain knowledge of them, (because they are not causal) or, alternatively, because our moral beliefs are subject to

Of course, putative cases of mathematical explanation are not without their critics. Some suggest that the mathematics is playing merely a representational role, and while the explanation may be presented in mathematical form, the real explanation lies in the physical phenomena in question (Melia 2000; Melia 2002). Others admit that the

---

debunking: given their causal history, it would be a massive fluke if, in fact, they tracked the moral facts (presumably in large part because if they are non-causal, then it's hard to see why causal mechanisms would reliably track such facts). (That is, moral beliefs are not *sensitive*). On this latter view our moral beliefs are unreliable, and hence not justified. This paper does not focus on this second worry. We do not, directly, attempt to undermine these kinds of debunking arguments. Nevertheless, some of what we say in this paper can be put to use in addressing such worries. As Clarke-Doane (2015) points out, some have thought that these debunking arguments do not apply to the mathematical case because mathematics is indispensable to our best explanations, in a way that moral properties are not. The idea is that we can gain mathematical knowledge through ordinary scientific knowledge via the role that mathematics plays in scientific explanation. Since moral properties are typically taken not to play any such explanatory role, this route to blocking debunking arguments is not available to the non-naturalist moral realist. If we are right, and even the non-naturalist can hold that moral properties are explanatory, then she, too, can respond to the debunking worry in the same way as the mathematical realist. Of course, this assumes that one thinks that the fact that mathematics is explanatory does in fact block the debunking argument. Not everyone accepts this. Clarke-Doane, for instance, (2015) argues that in general, showing that the contents of our D-beliefs figure in their best explanation is not sufficient for answering the 'reliability challenge' for D-realism. So, he argues, it's not the case that the moral realist is worse off than the mathematical realist in this regard. He argues that what we need to show is that, had the relevant truths been different, our beliefs (about those truths) would have been correspondingly different. Clarke-Doane (2015) goes on to try to show that this is so in the case of moral beliefs, even if the contents of those beliefs *don't* figure in their best explanation. By contrast, if what we say in this paper is correct, then moral and mathematical beliefs are sensitive in this manner—they would have been different, had things gone differently—and that is because of a counterfactual connection between the moral truths and moral beliefs.

mathematics is explanatory, but deny that this gives us reason to give ontological rights to mathematical entities (Leng 2010; Yablo 2005; Yablo 2012).

Our goal is not to argue that mathematics does play a genuinely explanatory role in science. That is a large task, beyond the scope of this paper. What matters is whether the moral non-naturalist is in a position to deny the explanatory power of mathematics. While we believe that a determined moral non-naturalist can endorse such a view, doing so is by no means straightforward.

In order to deny that mathematics plays a genuine explanatory role in science, one must show that mathematics is dispensable to scientific explanations. This requires doing two things. First, one must show how to reformulate our best scientific explanations without mathematics. Second, one must show that the reformulated explanations are at least as good as the originals with respect to explanatory power. That project, though it has been attempted (see Field 1980), has been widely held to fail (e.g. see Colyvan (2001: 67–90) and Malament (1982)).

In response, a moral non-naturalist might deny that there is any need to complete such a nominalisation project. She might, instead, draw upon the so-called easy road to nominalism in order to deny the explanatory criterion of ontological commitment. Easy-road nominalism involves accepting that mathematics plays an indispensable (though perhaps not explanatory) role in science while nonetheless denying that we should thereby be committed to the existence of mathematical entities. Some easy-road nominalists go further and accept that mathematics plays an indispensable explanatory role in science but deny that this commits us to mathematical realism. Such easy-road

nominalists thus deny the explanatory criterion of ontological commitment without reformulating our best scientific theories, and so may be a better crowd for the non-naturalist to fall in with (see, for instance, Leng (2010) and Azzouni (2004)).

The problem is that easy-road nominalists do not deny the explanatory criterion of ontological commitment in a manner that is helpful to the moral non-naturalist. Easy road nominalists concede the necessity of explanatory indispensability for ontological commitment but deny that explanatory indispensability is sufficient for ontological commitment. Rather, it is explanatory indispensability plus some further commitment that yields sufficiency. Easy road nominalism is no help to the moral non-naturalist, who wants to deny even the necessity of explanatory indispensability for ontological commitment. The only helpful way of severing the tie between ontological commitment and explanatory indispensability for the non-naturalist will retain ontological commitment without explanatory indispensability, not explanatory indispensability without ontological commitment.[9]

4. Moral Explanation Again

In the previous section, we considered a partners-in-crime response to the failure of non-naturalistic moral realism to provide moral explanations. We suggested that the non-naturalistic moral realist can no longer defend their view by appealing to the lack

---

[9] Thanks to a referee for this helpful way of making this point. A third option would be to follow Clarke-Doane (2014) and maintain that although some mathematics is indeed indispensable, while no moral facts are indispensable, in fact the majority of mathematics turns out *not* to be indispensable. Hence so long as we think that we have reason to believe in mathematics *in toto,* then we must think that we have reason to believe in many posits that play no indispensable role in explanation. If so, then the mathematics case does not look very different from the moral case.

of explanation within mathematics. That's because it is an open question as to whether mathematical entities are explanatorily dispensable. In this section, we offer an alternative strategy for defending non-naturalistic moral realism.

The idea is straightforward: the moral non-naturalist should—like the moral naturalist—maintain that moral properties *are* explanatory. As discussed, the reason moral non-naturalists tend not to go this way is because it is assumed that, since non-natural moral properties cannot do any causal work, they cannot do any explanatory work. It is this last inferential step that we believe the moral non-naturalist should resist. The moral non-naturalist ought to maintain that non-natural moral properties can do explanatory work, albeit of a non-causal kind.

In this way, the kind of explanation to which the moral non-naturalist appeals is of a piece with the kind of explanation to which the mathematical realist appeals. For whatever else one thinks of the explanation of physical facts by mathematical ones, the type of explanation implicated in the mathematical case is non-causal explanation. Mathematical facts about prime numbers do not cause North American cicadas to have prime-numbered life cycles. But those facts explain the prime-numbered life cycles in question nonetheless. Similarly, non-natural moral properties do not cause moral beliefs, but they explain those beliefs anyway.

We propose that the explanatory role played by non-natural moral properties is analogous to the explanatory role envisaged for mathematical facts. In both cases, the features in question make a difference to some physical phenomenon, in this sense: if the relevant (mathematical or moral) feature had not been thus and so, then the physical facts would not have been thus and so. In other words, we seek to understand the explanatory role played by non-natural moral properties in the same broad

counterfactual terms that are used to understand the explanatory role played by naturalistic moral properties. The difference, however, is that we do not seek to understand the counterfactual relationship in purely causal terms.

In response, one might argue that any counterfactual relationship that is genuinely explanatory must be a causal relationship, and so the imagined response on behalf of the non-naturalist is simply not viable, unless they are willing to posit there being non-natural moral properties with causal powers. In order to defend against this kind of worry it is instructive to return to the mathematical case. In the mathematical case, the explanatory contribution of mathematics can be analysed in counterfactual terms as well. But there is little temptation to treat the counterfactual relationships as indicating a causal relationship of any kind. As before, if this line works in the case of mathematics then it is a view that the moral non-naturalist can utilise in order to take full account of the explanatory power of non-natural moral properties.

Here's the idea. In response to scepticism about the idea that mathematical facts genuinely explain physical facts, mathematical realists have sought to defend the explanatory power of mathematics. The resources that mathematical realists have developed for this purpose are, we contend, available to non-naturalistic moral realists. This could be shown in a number of ways (depending on how, exactly, one conceives of mathematical explanations), but, as noted, we're going to focus on one strategy in particular, which appeals to recent work on counterfactuals.

One way to defend the existence of mathematical explanations is via the same counterfactual test for explanatory relevance that Sturgeon suggests for moral properties. However, in the mathematical case—and this will become important later on—the relevant counterfactuals have impossible antecedents (usually referred to as

'counterpossibles'). We can understand what it is for facts about mathematics to be explanatorily relevant to physical facts by considering what would have been the case had the mathematical facts been different. In order to gain a more concrete sense of how this works, it is useful to consider one case of mathematical explanation in a bit more detail. Consider the cicada case. As discussed, in the cicada case we are trying to explain why it is that cicadas have prime-numbered life-cycles. That explanation appeals to facts about predation, along with mathematical facts about how it is that primes factorise.

Now, consider the following counterfactual

CF1   If 13 had not been prime, then North American magicicadas would not
            have had 13-year life-cycles.

Baron, Colyvan and Ripley (2018) argue that it is possible to reason coherently about the truth of CF1. They defend this claim by appealing to a very general strategy for evaluating counterfactuals. This general strategy appeals to three notions that show up in one respect or another in all semantic theories of counterfactual conditionals. These are: holding fixed, twiddling, and ramifying. When evaluating a counterfactual, we first make a choice about what to hold fixed. These are the facts that we are keeping invariant under counterfactual variation. We then allow the facts corresponding to the antecedent of the counterfactual under consideration to vary: that is, we 'twiddle' the facts in order to consider a situation in which the antecedent is true. Third, we take note of the implications of the twiddle that we have made through the 'free facts': the facts that we are not holding fixed. If the ramification of the twiddle through the free facts vindicates the consequent of the counterfactual of interest, then the counterfactual is true; if not, then not.

So, for instance, consider the following familiar example. Suppose that one strikes a match and then the match lights. Now consider the following counterfactual:

CF2   If the match had not been struck, it would not have lit.

In order to evaluate this counterfactual, we first make a decision about what to hold fixed. In this case, we hold fixed the full history of the world up until the striking of the match. We also hold fixed as many laws of nature as we can compatible with twiddling the antecedent of the conditional. That is, we give up only whatever laws we need to give up in order to prevent the match from being struck, given that we are holding the history of the world fixed (that is, we introduce small 'miracles'). We then see what the implications are of the lack of a striking for the rest of the facts that we are not holding fixed. If the match lights anyway, then the counterfactual is false. If the match does not light, then the counterfactual is true.

Baron, Colyvan and Ripley (2018) apply the same reasoning to the mathematical case, with respect to CF1. Hold fixed the physical laws of nature, along with as many facts about number theory as you can, and hold fixed up to the point in the evolutionary history of the cicadas where they begin to be predated by predators with periodical life cycles. Next, and this is crucial, hold fixed a representation relation between the mathematical and physical facts. Thus, we hold fixed the fact that the life-cycle lengths of the cicadas are measured in years and are represented by the numbers 13 and 17 respectively. We do this so that when we consider a counterfactual alteration to the mathematical facts, the natural facts must bend and twist in response to the counterfactual alteration we are making, in virtue of the need to preserve the representation relation.

Next, we make a twiddle to the mathematics, by making 13 no-longer prime. [10] Intuitively, the change to 13's primeness would have implications for cicada life cycle length, namely: the cicadas would no longer "seek out" a 13 year life cycle, since 13 would no longer be optimal. Of course, the kind of twiddle that we are considering is an impossible one: it is necessarily the case that 13 is a prime number. Nonetheless, Baron, Colyvan and Ripley argue that good sense can be made of impossible perturbations to mathematical objects and, accordingly, that counterfactuals such as CF1 can be coherently entertained.

There is a great deal to say about the cicada case, and about the counterfactual approach taken by Baron, Colyvan and Ripley. For now, however, we will abstract away from the details of the case, and of the counterfactuals in order to make a broader point. If Baron, Colyvan and Ripley are correct, then one strategy for showing that mathematical facts are explanatory involves appealing to counterfactual variations: twiddle the mathematics and see if the physical facts twist and bend in response. If the physical facts do change in response to mathematical twiddling, then it is plausible to suppose that the mathematics is playing an explanatory role.

Notice, however, that the mathematical realist is not thereby committed to any claims about causation. Her view is not that, because there is counterfactual dependence of physical facts on mathematical facts, that there is thereby causal dependence. Rather, her view is to simply take the counterfactuals at face value, as telling us something

---

[10] There are a variety of ways to do this. The most straightforward is to imagine an impossible situation in which multiplication works differently to the way it works actually. Whereas actual multiplication is such that 13 has only itself and 1 as factors, an impossible counterpart of the multiplication relation, multiplication*, has it that 13 has 2 and 6 as factors as well.

about explanation that is independent of any causal facts. Our suggestion is that the non-naturalistic moral realist can defend the presence of moral explanations in the same way: by simply considering counterfactual variations to moral facts. The idea, in essence, is to apply Sturgeon's counterfactual test for explanatory relevance, while resisting the move to read this test in a causal spirit.

Our claim, then, is that the difference-making embedded in the relevant counterfactuals signals the presence of a non-causal style of explanation, and one that is analogous to the style of explanation offered by mathematical realists. If so, the non-naturalist can accept that moral properties are explanatory, just as easily as can the naturalist.

One might, at this point, disagree, responding that the naturalist is in a better position than the non-naturalist. This is because the non-naturalist faces a causal exclusion worry that the naturalist does not face. Both the naturalist and the non-naturalist maintain that the moral properties supervene on the natural properties, and that an observer's belief that (say) lighting a cat on fire is wrong, is caused by the subvenient natural properties, namely (say) the cat's suffering. The causal relationship between the subvenient natural properties and the observer's belief thus seems to screen off the supervenient moral properties. Even if the supervenient moral properties can do non-causal work *in principle*, in practice the subvenient natural properties provide a complete explanation that happens to be a causal explanation. So even if one is prepared to allow non-causal explanations into the picture, there is simply no explanatory work for the moral properties to do, since they play no role in the (complete) causal explanation of the situation.[11]

---

[11] See Kim (1989, 1998 and 2005) for the most thorough statement of the causal exclusion problem.

The first thing to notice is that this is not just a problem for the non-naturalist. The naturalist and the non-naturalist both face the same exclusion worry. Whether the moral properties are natural properties or non-natural properties, the subvenient natural facts threaten to render them explanatorily impotent. One might respond, of course, that the problem facing the non-naturalist is much more severe than the problem facing the naturalist. The naturalist can respond to the exclusion worry in ways that the non-naturalist can't. On the face of it, the naturalist has three responses available to her.

First, the naturalist can respond that there is token identity between the supervenient and subvenient properties.[12] If she takes this line, then she can straightforwardly accept that the moral properties do explanatory work: they do the very causal work that the natural properties do in bringing about an observer's moral beliefs. Second, the naturalist can suppose that the supervenient properties are grounded by, or ontologically depend on,[13] the subvenient properties;[14] so that the causal powers of the grounded are *inherited* from the ground, rather than in competition with the causal powers of the ground itself. The causal explanation can then be taken to consist in both the low-level natural properties in conjunction with the causal powers of the higher-level moral properties. Third, the naturalist might argue that the higher-level properties and the lower-level properties are both causally efficacious because in such cases there is causal overdetermination, but not overdetermination of an offensive or problematic kind. So,

---

[12] See Jackson (2012) for a recent defence of such a view. Also see Gozzano and Hill (2012) for an extended discussion of the identity theory.

[13] See Schaffer (2009) for an extended discussion of grounding and ontological dependence.

[14] Kroedel and Schulz (2016) suggest this option.

to be sure, the natural properties are sufficient for explaining a given physical phenomenon, but so too are the moral properties.[15]

In short, the naturalistic moral realist can argue that moral properties play an indispensable role in causation and thus in the total causal explanation for a given phenomenon. By contrast, the non-naturalistic moral realist cannot say any such thing. Non-naturalistic moral realism is the view that moral properties are *strongly distinct from* the natural properties: they cannot be identified with, nor do they reduce to, nor are they wholly grounded by, the natural properties. So the non-naturalistic moral realist cannot solve the causal exclusion problem by pointing to some intimate connection between the supervenient and subvenient properties (i.e. token identity, or some relation of grounding) whereby non-natural properties come to play a causal-explanatory role by inheriting the causal powers of the subvenient properties.

In fact, we suspect that the non-naturalistic moral realist *can* avail herself of the naturalist's responses to the causal exclusion worry. The non-naturalist thinks that there is some relationship between the moral properties and the natural properties: after all, she thinks that the former supervene on the latter, and that can't be a matter of mere fluke.[16] Whatever that connection consists in—even given that it's not token identity, or grounding—it is conceivable that the causal powers of the subvenient natural properties are inherited by the higher-level non-natural moral properties.

---

[15] For a general discussion of causal exclusion in the case of mental and physical properties see Andersen (2015).

[16] That being said, there are some non-naturalists who deny the supervenience of the moral on the natural. See, for instance, Roberts (2018) and Hattiangandi (2018). These non-naturalists may require the same kind of picture as the 'strong' non-naturalism we describe in the next section.

Suppose, however, that the non-naturalist cannot make use of something like the naturalist's solution to the causal exclusion problem. Even then, we don't yet see a reason to give up on the explanatory power of the moral facts.

To see why, note that a similar problem arises in the mathematical case. Consider the cicadas once again. One might argue that there is a range of physical facts that the mathematics represents which are sufficient for bringing about, causally, the prime-numbered life-cycle lengths of the cicadas. These include causal facts about cicada breeding, as well as causal facts involving predation. Indeed, it is plausible that the evolutionary history of the cicadas is a large, complicated network of causal relations extended through time. It is also plausible that the causal facts are causally sufficient for the cicadas to have the life-cycles that they do. One might go on to argue that the causal facts explanatorily exclude the mathematical facts, in the following sense: in each putative case of mathematical explanation, there is a sufficient causal explanation available that, as a matter of fact, does all of the explanatory work. Thus, even granting that the mathematics can play a role in non-causal explanation, there is simply no explanatory work left for the mathematics to do. Clearly, the moral naturalist's responses to the exclusion argument won't work here. Mathematical facts do not stand in the right kind of relation to inherit causal powers from the natural facts that they represent. So, if the complete explanation for a given physical phenomenon is a causal explanation, then, since the mathematical facts—like the moral facts—are causally screened off, they fail to be explanatory.

The solution, in the mathematical case, is to deny that the causal explanation is sufficient in any given case. Rather, the mathematics *adds* something to a given explanation such that, without the mathematical facts, the explanation would be impoverished in some respect. In each case, the causal facts alone are not sufficient to

explain whatever it is we are trying to explain. The counterfactuals that relate the mathematical and the physical are evidence of the explanatory work that the mathematics can do, but should not be interpreted in a causal spirit. If that's the right solution in the mathematical case, then the same solution can be applied to the case of moral non-naturalism. The moral non-naturalist can admit that the moral properties do no causal work, while maintaining that they do explanatory work nonetheless. She can then draw on her partner in crime to defend her position, maintaining that if the counterfactuals are evidence of a non-causal explanatory contribution in the mathematical case, then the same is true in the moral case. If one denies the capacity of the moral facts to explain, then one must deny the capacity of the mathematical facts to do similar work.

Of course, the question that then arises in both the moral and the mathematical cases is: what is the explanatory contribution made by the mathematical or moral facts? If one is prepared to deny that the only way for something to be explanatory is for it to do causal work, then one owes some account of what the explanatory contribution of the moral and mathematical facts might be. As before, this is not a problem that the moral naturalist faces. She can simply say that the explanatory contribution that the moral properties make is a causal contribution. Once again, however, the moral non-naturalist can lean on the mathematical realist's response to this question. The mathematical realist maintains that the mathematics contributes something to the explanation of certain physical facts that no causal explanation can deliver.

Thus far there is no agreed-upon account of what non-causal explanation is or how it works.[17] There are, however, two common themes in philosophical thinking about

---

[17] See Sober (1983) and Baron and Colyvan (2016) for discussion.

such explanations. The first is that in such explanations, things couldn't have been otherwise (for some relevant modality). Typically, there are causal processes in place, but they are irrelevant in the sense that, no matter how the causal facts turned out, the result would stand and it is the non-causal explanation that delivers this modally robust result (Colyvan 2012, Pexton 2016).[18]

The second theme is that non-causal explanations deliver a variety of explanatory unification that cannot be had otherwise. Baker (2017), for instance, argues that in the mathematical case, the higher-level mathematical facts allow us to unify a range of explananda under a common explanatory scheme. No purely causal facts can provide any such explanatory power. That's because, in the cases he considers, the causal facts are too different across the various cases of explanation to be plausibly thought to constitute a unified explanation. Baker argues that the cicada case is like this. He shows how to unify the cicada case with a range of other cases involving gear ratios in mechanical systems. The causal facts across the various cases vary massively: each case involves a different group of physical laws, and a different type of causal system. Nonetheless, mathematically the cases are all unified: the same properties of prime-

---

[18] We have focused on mathematical explanations of physical facts, but there are other non-causal explanations that appear to be modally robust in a similar manner. For instance, Baron and Colyvan (2016) argue that the law of non-contradiction explains why a time traveller travelling to their local past cannot kill their grandfather. They maintain, further, that no causal explanation can do this work. Still others involve mathematics to explain within mathematics: some, but not all, mathematical proofs of a particular theorem show more than *that* the theorem is true; they explain *why* it is true (Lange 2017 and Colyvan et al. 2018, Steiner 1978). In the last case (of a mathematical proof explaining a mathematical theorem) there are no causal processes at all; there is only mathematics.

numbers that explain the life-cycle lengths of North American cicadas also explain why it is that fixed-speed bicycles are produced with gears that possess a prime number of teeth, and thus rotate according to a prime-numbered cycle.

Both the modal robustness of explanation and the unifying power of explanation are potentially available to the moral non-naturalist. Sturgeon (1985) outlines an example of moral explanation that appeals to the way in which a government is unjust. The thought is that the government being unjust explains why there is eventual rebellion against the government. The fact that the government is unjust is identified with various moral properties of the agents running the government (or, indeed, of the government itself, if there are group agents with moral properties).

Now, clearly, there are many different ways for a government to be unjust, and there are many different ways that this can result in rebellion. It seems plausible to suppose, however, that for at least some ways of setting up an unjust government, the initial set-up of the government will eventually lead to open rebellion. Of course, this is not going to be true in every case. Nonetheless, if the government is both sufficiently morally egregious and sufficiently weak to open revolution by a populace that is, itself, well-equipped, then rebellion is very likely indeed. The causal facts on their own in any given case of rebellion don't explain why, *for other governments*, the same outcome has to occur. Moreover, the moral facts seem to be required to explain the consistency of the outcome. Certain modally strengthened counterfactuals can help us to see this; strengthened in the sense that the consequent is some modally robust state of affairs. For instance, it seems at least defensible to say that if the government had not been unjust *for a certain kind of state*, then rebellion would not have been inevitable.

We admit that the modal contribution of moral properties is controversial. Still, the point remains that in order to continue pressing the partners-in-crime defence, the non-naturalist would be well served by showing that their moral explanations share the modal robustness (for some, though not necessarily the same, modality) of mathematical explanations.

The unifying potential of moral explanations appears even more promising. Consider that many observers form a morally relevantly similar moral belief in response to a wide-range of different events. The relevantly similar moral belief, in each case, is that the relevant action is morally wrong. Our imagined observer of cat immolation forms the belief that cat immolation is wrong, while most of those who hear of Hitler's genocidal actions form the belief that those actions were wrong. Similarly, most form the relevantly similar belief in response to the actions of an unjust government: namely that *those* actions are wrong.

Of course, there are differences between the beliefs: each is a belief that a different action is wrong. But there is an overall similarity to the beliefs as well, and that similarity lies in the unified moral response that agents have to a range of different actions. This similarity in what we are trying to explain—the moral beliefs of observers—cannot be accounted for in purely causal terms. The causal pathways that bring about the various moral beliefs are all quite different. Indeed, some such causal pathways may have virtually nothing in common with one another. They may be as different as murder and theft. Nonetheless, the similarity is unmistakeable. The moral facts can explain this similarity. We can unify a range of different moral responses together by appealing to the fact that all of the causal facts carry a particular moral

property, and this moral property non-causally explains the similarity in the moral beliefs of agents.[19]

5. The Moral Law

So far, we have argued that the moral non-naturalist has the same capacity to accommodate moral explanations as the moral naturalist. In this section, we clarify the scope of this defence of the explanatory power of non-natural moral facts. In particular,

---

[19] Skeptics about the explanatory indispensability of moral facts do not always deny that such facts might be indispensable to explanations that are of value in a pragmatic sense---e.g., explanations that would satisfy a certain audience, or help them achieve understanding about a certain matter. Their complaint is, rather, that moral facts do not seem indispensable to the best available explanations that are valuable in an epistemic sense: a sense that is truth-tracking, not merely pragmatic, and which supports ontological commitment (Harman 1977: Ch.1; Shafer-Landau 2003: 100f; Woods 2018: 54f). Above, we suggested two respects in which explanations citing non-natural moral facts are arguably superior to explanations citing only natural facts: the former make for explanations that have the right sort of modal robustness, and that are more unified than the latter. We believe that these virtues of non-naturalistic moral explanations are not merely pragmatic. Rather, they are epistemic virtues and are therefore of the sort that support metaphysical commitment. Our basis for this claim stems from the importance of both unification and modal robustness in the burgeoning literature on scientific explanation. Both unification and modal robustness are seen to be epistemic virtues of explanations, and are often used to ground theories of explanation more generally (for the unification case, see Kitcher (1981); for the modal robustness case, see Woodward (2003)). We admit, however, that we cannot settle the issue here. There is plenty of scope to disagree about which virtues of an explanation are merely pragmatic virtues, and which are epistemic, and about whether there is any clear distinction to be drawn between these two notions at all. There is a much broader debate to be had about what counts as a good explanation.

we will argue that the defence only works for what we will call *weak* non-naturalism; it doesn't work for a strengthened version of the view.

First, it is useful to draw a distinction between moral properties and moral laws. Moral properties are properties such as Hitler's depravity cited above. These are properties possessed by agents, actions, beliefs and so on. The moral laws encode the relationships between various moral properties, and between moral properties and natural properties. According to *weak* non-naturalism, the moral laws are just a systematisation of the moral properties. The moral laws thus depend upon the moral properties, which in turn supervene on the natural properties. Moral properties are not grounded in the moral laws, in any meaningful sense.

According to what we will call *strong* non-naturalism, the moral laws are not merely a way of systematising the moral properties; they are independently existing moral facts that govern the way that the moral properties work, and the way that they interact with the natural properties. On this view, the non-natural moral properties are at least partly grounded in the moral laws. The moral properties still supervene on the natural facts, however. Indeed, their supervenience is partly encoded by the moral laws that dictate the relationship between the moral and the natural properties.

On both views, the moral laws are assumed to be necessary truths, so they trivially supervene on the natural properties. Hence both weak and strong naturalists agree that any two *possible* worlds[20] that are alike with respect to their distribution of naturalistic

---

[20] Here, we take 'possible worlds' to simply be the broadest sphere of genuinely possible worlds: what are also sometimes known as the metaphysically possible worlds. Any world that is not within this sphere is, for us, an impossible world. Whether all impossible worlds are therefore worlds that contain a contradiction, or whether there are impossible worlds that are in *some* sense internally coherent, is not a

properties are alike with respect to their moral law and, in turn, alike with respect to their moral properties, though they disagree about why this is so.

Weak non-naturalists can provide moral explanations in the manner described in the previous section. So, for example, consider the following counterfactual:

CF3   If lighting the cat on fire for amusement had not been morally wrong, then the observer of the teens' action would not have formed the belief that lighting the cat on fire for amusement is morally wrong.

Just as the mathematical realist can defend the existence of mathematical explanation by defending the truth of CF1, so too can the weak non-naturalist defend the existence of non-naturalistic moral explanations by defending the truth of CF3. To evaluate CF3 we use the same broad procedure described above both for ordinary counterfactuals and for counterfactuals involving mathematics. First, hold fixed the physical laws. Then, hold fixed all of the physical facts up until just before the cat is immolated. We also hold fixed the fact that the moral facts supervene on the natural facts. We do this for the same reason that we hold fixed the representation relation in the mathematical case. We do it so that when we make a counterfactual alteration to the moral facts, the subvening natural facts must bend and twist in order to preserve supervenience.

Now twiddle the moral facts, by making it the case that causing pain for amusement is not morally wrong, and consider the implications of doing so for the observer's belief formation. When we do this, it seems plausible that the observer would not have formed the belief that lighting the cat on fire is morally wrong. That's because, if we make it the case that lighting a cat on fire is not morally wrong, then we must alter the

_____

matter we will address here, since it makes relatively little difference to one's account of counterpossibility and counterpossibles.

underlying naturalistic facts. If the relevant natural fact is the cat's suffering, then the counterfactual alteration to the moral fact must result in the naturalistic fact being different. The cat cannot be suffering in the imagined scenario. But if the cat is not suffering, then the observer shouldn't form the belief that lighting the cat on fire is morally wrong (Sturgeon 1985: 66).

Despite the apparent parity with mathematical realism, for the weak non-naturalist CF3 is disanalogous to a mathematical counterfactual like CF1 in a certain respect. When considering mathematical counterfactuals, we need to consider counterpossibles: cases in which mathematical facts are different. When considering the relevant moral counterfactuals, as a weak non-naturalist, the counterfactuals need not be counterpossibles. That's because there is nothing impossible about a situation in which lighting the cat on fire for amusement is not morally wrong, so long as we allow the natural facts to vary. For we can simply set up a possible situation in which the cat doesn't suffer, is not having its life prematurely ended, is not losing utility and so on. In short, we can set the natural facts to be in a possible state that corresponds to the alteration to the moral facts that we have made.

When evaluating CF3 under strong non-naturalism, the kinds of counterfactuals we must consider may be more closely aligned with mathematical counterfactuals. It really depends on what the moral law says. Suppose, for instance, that the moral law says that lighting cats on fire for amusement is always, everywhere wrong come what may (this is like Kant's prohibition on lying or the supernaturalist's prohibition on killing). If the moral laws are necessary truths, then it doesn't matter how we set up the natural facts: there is no way to set them up so that morality gives the thumbs up to lighting the cat on fire. The only way to get the relevant difference in the moral facts is to alter the

moral law, and there is no *possible* way to alter the natural facts, that entails (or indeed, is consistent with) altering the moral law.

The moral law is, in such a case, analogous to mathematical facts. Altering the moral facts by altering the moral law requires the consideration of counterpossible scenarios. Now, on the face of it, this is not a problem. The machinery that the mathematical realist uses to model mathematical explanations can be used to handle moral explanations as well. Thus, just as a mathematical realist can think in terms of mathematical counterpossibles, so too can the moral realist think in terms of moral counterpossibles. These moral counterpossibles will be ones in which we counterpossibly twiddle the moral laws, and then look to see how things are with the moral facts.

Matters, however, are not so straightforward. As discussed, the mathematical realist makes use of a representation relation that obtains between the mathematical and physical facts. This representation relation is not a modal relation. Which means that it is not a relation that describes the relationship between two things in some set of possible worlds. It is a relation, rather, that can describe the relationship between two things in possible and impossible situations equally. Recall also that the non-naturalist's supervenience relation is supposed to play the same role as the mathematical realist's representation relation. In both cases, the relevant relation is held fixed so that when we counterfactually vary the moral or the mathematical, we get a corresponding change to the natural facts.

Now, the trouble is that supervenience, unlike representation, *is* a modal relation. Supervenience describes the relationship between two things in possible worlds only. Suppose, then, that we hold the supervenience of the moral facts on the natural facts fixed, and suppose that we consider an impossible scenario in which the moral law is

different. All that holding fixed the supervenience relation guarantees is that the moral and natural facts in the relevant world cannot come apart in any possible situation. But the relevant world is *not* a possible situation. And so, it is compatible with holding the supervenience relation fixed that the moral facts and the natural facts come apart. The upshot is that, even if we hold fixed the supervenience relation between the natural facts, the moral laws, and the moral facts, it doesn't follow that when we make a counterpossible change to the moral law, that there will be a corresponding change to the natural facts. There could be such a change. But equally, there might not be. The supervenience relation does not constrain the relationship between the moral and the natural in the impossible situation we are considering.

In short, we can't do the same work by holding fixed the supervenience relation in the moral case, which we can do by holding fixed the representation relation in the mathematical case. For all we have said, the strong non-naturalist has no way to establish the truth of CF3.

There is a potential way forward for the strong non-naturalist but, as we shall see, it comes with a cost. Instead of holding fixed the supervenience of the moral facts M on the natural facts N, we hold fixed the subvening properties for M—namely N—and consider a direct counterfactual relationship between M and the observer's belief P. Thus, we say that M explains P directly because, holding N fixed, if M had not been the case, then P would not have been the case.

To see the difference, we can contrast two versions of CF3. According to the weak non-naturalist, CF3 involves probing the explanatory potential of moral facts by looking at counterfactual variations to natural facts. By contrast, for the strong non-naturalist, CF3 involves probing the explanatory potential of moral facts by looking at counterfactual

variations to moral facts, given that the natural facts are thus and so. The first kind of counterfactual seems appropriate if moral facts supervene on natural facts, with no intermediary via a moral law.

If, however, moral facts are grounded in the moral law, (which in turn supervenes on the natural facts) then it is odd to use the first interpretation of CF3 to investigate the explanatory importance of the moral facts. In this situation, the second interpretation of CF3 seems more appropriate, since it does not commit one to looking at the explanatory contribution of the moral facts as some function of the natural facts. All we need do is twiddle the moral facts directly, by, say twiddling the moral law, and then seeing what happens to the observer's beliefs. The natural facts can simply be held fixed, since their explanatory contribution, such as it is, does not enter into the picture.

But why think that CF3, thus understood, is true? We are not certain that it is. However, preliminary evidence for CF3 can be given via the following thought experiment. Suppose that a brand of consequentialism is true in the actual world. Now, imagine a world that is just like ours with respect to all of the natural facts, but in which consequentialism is false. In that world, a peculiar deontic theory is true, according to which causing pain to animals is not morally wrong, because only persons are subject to moral consideration. Animals, we can imagine in this world, have the moral status of inanimate objects. Now suppose that, in that world, a group of teens light a cat on fire, and an observer witnesses the lighting of the cat. In this situation, we are to imagine that the observer would not form the belief that the lighting was morally wrong (we return to this contention momentarily). *If* that's right, then we have a reason to suppose that CF3—or something like it—as interpreted by the strong non-naturalist is true.

If CF3 so interpreted is true, then the strong moral non-naturalist can allow that her non-natural properties are directly explanatory. On the face of it, this brings her view into parity with the mathematical realism discussed in §4.

There is, however, a substantial difference between the two positions, which means that strong non-naturalism is at a relative disadvantage to weak non-naturalism when it comes to depending on the analogy with mathematical realism.

Let's return to the contention that the observer would not form the belief that the lighting was morally wrong. When we hold fixed the natural facts (in particular the fact that the cat is lit on fire) and change only the moral law, and the moral fact that is grounded in that law, we leave in place all of the causal facts that are relevant with respect to belief formation. In short, by holding the natural facts fixed and changing only the moral facts, we will find no difference to the observer's belief. If that is right, then CF3 is false: counterpossibly changing the moral facts does not result in a change in the observer's beliefs

Such a view assumes that the total explanation for belief formation is a causal explanation, and so, by holding fixed the causal facts, we have thereby held fixed everything that makes an explanatory difference to the observer's belief. Should the strong non-naturalist accept that this is so? Notice that the strong non-naturalist[21] is *already* committed to the view that the moral beliefs of agents somehow track the moral law, and by so doing, track the moral facts. Exactly how we successfully gain knowledge of the moral law, given that it is not grounded in, or reducible to, any natural facts, is one of the outstanding challenges for the strong moral non-naturalist. It seems to be a necessary condition on meeting any such challenge that our moral beliefs are

---

[21] At least of a non-skeptical persuasion.

shown to counterfactually depend on the moral law, in this sense: if the moral law had been different, our moral beliefs would have been different. Given this assumption, a strong non-naturalist may well deny that when holding fixed the natural facts we have already held fixed everything that makes an explanatory difference to our beliefs. If, for instance, she thinks that we gain moral knowledge by directly apprehending the moral law, then she is already committed to the view that the natural facts *alone* don't account for our beliefs; there is some further role for the moral law (and hence moral facts) to play.

If we assume that the strong non-naturalist has a viable epistemology of the moral law, then one cannot suppose that the natural facts entirely determine an observer's belief when we are evaluating CF3. Rather, one has to allow that an observer's belief changes in line with changes in the moral law, quite independently of the natural facts. This effectively collapses the question of whether the strong non-naturalist can provide moral explanations into the question of whether she has a viable moral epistemology. This is beneficial insofar as it reduces the number of outstanding issues facing the strong non-naturalist. It is problematic insofar as it weakens the analogy with mathematical realism.

Mathematical realists, like strong non-naturalists, face a difficulty when it comes to providing a reasonable epistemology. Mathematical objects are not located anywhere in space or time, and so it is unclear how we could have any epistemic access to them. Some mathematical realists maintain that the epistemic problem for mathematical realism can be solved via the provision of mathematical explanations. In particular, if it can be shown that mathematical facts explain physical facts, then we can gain knowledge of mathematical objects in the same way that we gain knowledge of

unobservable entities in science more generally: through the use of inference to the best explanation.

The strong non-naturalist cannot approach the epistemic problem that they face in the same manner. If providing moral explanations is a matter of providing direct explanations of moral belief formation—explanations that are not mediated by natural facts—then one cannot hope to solve the epistemic problem by first securing moral explanations, since solving the epistemic problem just is a matter of providing the relevant explanations. Thus, if in order to establish the truth of CF3 one must assume an adequate moral epistemology, then one clearly cannot use CF3 to produce a viable moral epistemology.

Despite the fact that the strong non-naturalist and the mathematical realist fail to share the same dialectical situation, the strong non-naturalist is not completely on their own. They *can* appeal to the same machinery as the mathematical realist to make sense of counterpossibles involving moral facts. They still, however, owe an account of how we come to know the moral laws, given that it is not via any knowledge of the natural facts.

Of course, one might seek to take issue with CF3 on the very grounds that it is a counterpossible. According to some philosophers, all counterpossibles are (at best) trivially true. Indeed, this is a straightforward implication of the standard Lewis-Stalnaker semantics for counterfactuals (Lewis (1973) and Stalnaker (1968). Thus, the naturalist might argue, while CF3 is true, so is:

CF3* If causing pain for amusement had not been morally wrong, then the observer of the teens' action would have formed the belief that lighting the cat on fire for amusement is morally wrong.

But if CF3* is true as well, then the explanatory contribution of the non-natural moral properties at issue appears to be wiped out. That's because there are nearby worlds where the person retains the belief and there are nearby worlds where the person lacks the belief.

There are two broad responses to this worry about counterpossibles. First, there exists a straightforward extension of the Lewis-Stalnaker semantics for counterfactuals that does not trivialise in this fashion.[22] The basic idea is to extend the semantics so that it ranges over both possible and impossible worlds. Given the existence of these alternative semantic pictures, the naturalist owes an argument in favour of holding on to the Lewis-Stalnaker semantics.

Second, and more importantly, even the naturalist should concede that counterpossibles involving moral facts are sometimes true in a non-trivial way. This is so for two reasons. First, it is a part of common sense that some counterpossibles involving moral facts are true. These counterpossibles need not be counterpossibles like CF3. Rather, consider a more ordinary counterfactual like: if allowing people to suffer is morally obligatory, then we morally ought to dismantle national health care. Now, compare this with the following: if allowing people to suffer is morally obligatory then we morally ought not dismantle national health care. Intuitively, the first counterfactual is true, while the second counterfactual is false. If allowing people to suffer is morally obligatory, then this is a necessary truth: there is no possible world in which allowing people to suffer is not morally obligatory. So, both counterfactuals are counterpossibles. If, however, we allow that counterpossibles such as these are non-trivially true (or false) then one cannot take umbrage with CF3 on the grounds that it too is a counterpossible.

---

[22] See Mares (1997), Nolan (1997), and Priest (2002).

Second, ethical debates over what the correct first-order moral theory might be are up to their eyeballs in counterpossibles. For instance, consider the following counterfactual: if act-utilitarianism were true, then it would be morally obligatory for the surgeon to secretly kill the one innocent patient and distribute his organs to five others.[23] This kind of counterfactual is a standard tool for testing the viability of moral theories, and in probing the consequences of adopting this or that theory of normative ethics. However, if act-utilitarianism is false, it is necessarily false. It is not the case that in our world, some version of deontology is the correct moral theory, while in some far-flung possibility, utilitarianism is correct. Given this, the above counterfactual is a counterpossible. Accordingly, at least some of these counterfactuals are going to be counterpossibles, depending on what the correct moral theory turns out to be (if there is one). [24]

---

[23] Suppose that the usual stipulations hold in this case: e.g., the patient's organs are a match; the lives of the five are sufficiently happy to outweigh the loss of happiness of the one, etc.

[24] One might wonder if the naturalist can accommodate these kinds of first-order disputes without appealing to impossible worlds. A referee suggested that one could evaluate these claims by appealing to what an ideal observer who endorsed some particular moral theory, would say in a possible world. Here, we think it depends what one intends by an 'ideal observer'. If ideal observers are epistemically ideal—if they can reason faultless—then on at least some accounts of possibility/impossibility, such observers will not endorse an impossible theory: for what it is to endorse an impossible theory is to be in some way epistemically limited, such that correct reasoning does not reveal that that theory is impossible. Nevertheless, we concede that there might be ways to make sense of these claims that do analyse them counterpossibles, and also, that there are ways to model counterpossibles, which do not appeal to impossible worlds.

Thus, if the naturalist maintains that all counterpossibles are trivially true, she runs the risk of undermining standard ethical theorising.

6. Conclusion

Our aim, in this paper, has been to show that the non-naturalistic moral realist has the resources available to provide moral explanations. This, in turn, furnishes the non-naturalist with a strategy for accepting the explanatory criterion of ontological commitment. If, as mathematical realists maintain, mathematics is genuinely explanatory, then what we have sketched is also a new partners-in-crime defence. Of course, for the mathematical realist who believes in mathematical explanation *there is no crime.* Looked at from this perspective, the parties are partners in a non-standard explanatory strategy.

References

Andersen, H. 2015. 'Mental Causation'. In N. Levy, and J. Clausen (ed.), *Springer Handbook of Neuroethics.* Dordrecht: Springer.

Azzouni, J. 2004. *Deflating Existential Commitment: A Case for Nominalism*, Oxford: Oxford University Press.

Baker, A. 2005. 'Are There Genuine Mathematical Explanations of Physical Phenomena?', *Mind*, 114(454): 223–238.

Baker, A. 2009. 'Mathematical Explanation in Science', *British Journal for the Philosophy of Science*, 60(3): 611–633.

Baron, S. 2014. 'Optimization and Mathematical Explanation: Doing the Lévy Walk', *Synthese*, 191(3): 459–479.

Baron, S. and Colyvan, M. 2016. 'Time Enough for Explanation', *Journal of Philosophy*, 113(2): 61–88.

Baron, S., Colyvan, M., and Ripley, D. 2017. 'How Mathematics Can Make a Difference', *Philosophers' Imprint*, 17(3): 1–19.

Benacerraf, P. 1973. 'Mathematical Truth', reprinted in P. Benacerraf, and H. Putnam (eds.), *Philosophy of Mathematics Selected Readings*, second edition, Cambridge: Cambridge University Press: 403–420. (First published in 1973.)

Brink, D. O. 1989. *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.

Clarke-Doane, J. 2014. 'Moral Epistemology: The Mathematics Analogy', *Noûs*, 48(2): 238–255.

Clarke-Doane, J. 2015. 'Justification and Explanation in Mathematics and Morality' In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics: Volume 10*, Oxford: Oxford University Press, pp. 80–103.

Colyvan, M. 2001. *The Indispensability of Mathematics*, Oxford: Oxford University Press.

Colyvan, M. 2002. 'Mathematics and Aesthetic Considerations in Science', *Mind*, 111(441): 69–74.

Colyvan, M. 2007. 'Mathematical Recreation Versus Mathematical Knowledge', in M. Leng, A. Paseau, and M. Potter (eds.), *Mathematical Knowledge*, Oxford: Oxford University Press, pp. 109–122.

Colyvan, M. 2010. 'There is No Easy Road to Nominalism', *Mind*, 119(474): 285–306.

Colyvan, M. 2012. 'Road Work Ahead: Heavy Machinery on the Easy Road', *Mind*, 121(484): 1031–46.

Copp, D. 2003. 'Why Naturalism?', *Ethical Theory and Moral Practice* 6(2): 179–200.

Field, H. 1989. *Realism, Mathematics and Modality*, Oxford: Blackwell.

Field, H. 1980. *Science Without Number*s, Oxford: Oxford University Press.

Gozzano, S. and Hill, C. S. (eds.). 2012. *New Perspectives on Type Identity: The Mental and the Physical*, Cambridge: Cambridge University Press.

Harman, G. 1977, *The Nature of Morality*, New York: Oxford University Press.

Hattiangandi, A. 2018. 'Moral Supervenience', *Canadian Journal of Philosophy*, 48(3–4): 592–615.

Jackson, F. 2012. 'In Defence of the Identity Theory Mark I.' in S. Gozzano, and C. S. Hill (eds.), *New Perspectives on Type Identity: The Mental and the Physical*, Cambridge: Cambridge University Press.

Kim, J. 1989. 'The Myth of Non-reductive Materialism'. *Proceedings and Addresses of the American Philosophical Association*, 63(3): 31–47.

Kim, J. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation,* Cambridge MA: MIT Press.

Kim, J. 2005. *Physicalism, or Something Near Enough*, Princeton: Princeton University Press.

Kitcher, P. 1981 'Explanatory Unification', *Philosophy of Science*, 48(4), 507–531.

Kroedel, T. and Schulz, M. 2016. 'Grounding Mental Causation', *Synthese*, 193(6): 1909–1923.

Lehman-Ziebarth, N., Heideman, P. P., Shapiro, R. A., Stoddart, S.L., Hsiao, C. C. L., Stephenson, G. R., Milewski, P. A., and Ives, A. R. 2005. 'Evolution of Periodicity in Periodical Cicadas', *Ecology*, 86(12): 3200–3211.

Lange, M. 2017. *Because Without Cause: Non-causal Explanations in Science and Mathematics*, Oxford: Oxford University Press.

Leng, M. 2010. *Mathematics and Reality*, Oxford: Oxford University Press.

Leng, M. 2016. 'Taking Morality Mathematically: Enoch's Indispensability Argument', in U. D. Leibowitz and N. Sinclair (eds.), *Explanation in Ethics and Mathematics: Debunking and Dispensability*, Oxford: Oxford University Press, pp. 204–219.

Lewis, D. 1973. *Counterfactuals*, Oxford: Blackwell.

Lyon, A. 2012. 'Mathematical Explanations of Empirical Facts, and Mathematical Realism', *Australasian Journal of Philosophy*, 90(3): 559–578.

Lyon, A. and Colyvan, M. 2008. 'The Explanatory Power of Phase Spaces', *Philosophia Mathematica*, 16(2): 227–243.

Malament, D. 1982. 'Review of Field's *Science without Numbers*', *Journal of Philosophy*, 79(9): 523–534.

Mares, E. D. 1997. 'Who's Afraid of Impossible Worlds?', *Notre Dame Journal of Formal Logic,* 38(4): 516–526.

Melia, J. 2000. 'Weaseling Away the Indispensability Argument', *Mind*, 109(435): 455–480.

Melia, J. 2002. 'Reply to Colyvan', *Mind*, 111(441): 75–79.

Miller, A. 2013. *Contemporary Metaethics*, 2nd ed., Malden: Polity Press.

Nolan, D. 2001. 'Impossible Worlds: A Modest Approach', *Notre Dame Journal of Formal Logic*, 38(4): 535–572.

Parfit, D. 2011. *On What Matters, Volume 2*, Oxford: Oxford University Press.

Pexton, M. 2016. 'There Are Non-Causal Explanations of Particular Events', *Metaphilosophy*, 47(2): 264–282.

Priest, G. 2002. *Beyond the Limits of Thought,* Oxford: Oxford University Press.

Putnam, H. 1971. *Philosophy of Logic*, New York: Harper and Row.

Putnam, H. 2004. *Ethics Without Ontology*, Cambridge: Harvard University Press.

Railton, P. 1989. 'Naturalism and Prescriptivity', *Social Philosophy and Policy*, 7(1): 151–174.

Roberts, D. 2018. 'Why Believe in Moral Supervenience?' in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Oxford: Oxford University Press.

Scanlon, T. M. 1998. *What We Owe to Each Other*, Cambridge: Harvard University Press.

Scanlon, T.M. 2014. *Being Realistic About Reasons*, Oxford: Oxford University Press.

Schaffer, J. 2009. 'On What Grounds What', in D. Manley, D. Chalmers and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Oxford University Press, pp. 347–383.

Sober, E. 1983. 'Equilibrium Explanation', *Philosophical Studies*, 43(2): 201–210.

Sturgeon, N.L. 1985. 'Moral Explanations', in D. Copps and D. Zimmerman (eds.), *Morality Reason and Truth: New Essays on the Foundations of Ethics*, Totawa, N.J.: Rowman and Littlefield, pp. 49–78.

Sturgeon, N. L. 2003. 'Moore on Ethical Naturalism', *Ethics*, 113(3): 528–556.

Shafer-Landau, R. 2003. *Moral Realism: A Defence*, Oxford: Oxford University Press.

Stalnaker, R. C. 1968. 'A Theory of Conditionals', in N. Rescher (ed.), *Studies in Logical Theory*, Oxford: Blackwell, pp. 98–112.

Steiner, M. 1978. 'Mathematical Explanation', *Philosophical Studies*, 34(2): 135–151.

Thomson, J. J. 1996. 'Moral Objectivity', in G. Harman and J. J. Thomson (eds.), *Moral Relativism and Moral Objectivity*, Oxford: Blackwell, pp. 65–154.

Woodward, L. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Yablo, S. 2005. 'The Myth of the Seven', in M. Kalderon (ed.), *Fictionalism in Metaphysics*, Oxford: Oxford University Press. pp. 90–115.

Yablo, S. 2012. 'Explanation, Extrapolation, and Existence', *Mind*, 121(484): 1007–1029.