Explanationism vs Modalism in Debunking (and Theory Choice)

Harjit Bhogal

Consider this line of thought: There seems to be a striking correlation between our moral beliefs and the moral truths. Given a non-naturalist realist conception of morality, where moral facts are mind-independent and causally inert, this correlation seems to be a coincidence. So, we have reason to reject the combination of non-naturalist realism and the correlation between our moral beliefs and the moral truth. This line of thought is at the heart of many formulations of *debunking* arguments against non-naturalism.

There are two main strands of response in the literature. Firstly, *third-factor* responses claim, roughly, that even given non-naturalist realism there is a common explainer of the moral beliefs and moral truths so the correlation between them isn't coincidental (e.g. Enoch, 2011; Skarsaune, 2011; Wielenberg, 2010, among many others).[1] Let's put this response aside to focus on a different one.

Perhaps the most influential type of response in the recent literature, and what I'll focus on, are *Modalist* responses. Such responses accept that there is no explanation of the correlation — there is no explanatory connection between belief and truth on the non-naturalist view. But, they argue, the correlation between moral belief and moral truth has certain important modal features which make it unproblematic. In particular, it's common to emphasize that our moral beliefs are *safe* and *sensitive* — and what more could we want from our moral beliefs than that? (Justin Clarke-Doane has been most influential in developing this position. See, for example, Clarke-Doane (2016a,b, 2020) and

[1] As a reviewer points out, Field (1996) first discussed, but rejected, this response.

Clarke-Doane and Baras (2019). Though, as Schafer (2017) and Faraci (2018) note, ideas in this modalist spirit run through a lot of the literature on debunking.)[2]

Explanationists, though, reject this modalist position. They claim that we want more than safety and sensitivity – we want an explanatory connection between belief and truth. Even if the correlation between belief and truth could not easily be false and even if our moral beliefs are safe and sensitive then the lack of such an explanation undermines our beliefs. (See, for example, Korman and Locke (2020), Lutz (2020), Faraci (2019).)

There's a standoff here. Explanationists claim that given non-naturalist realism our moral beliefs are explanatorily disconnected from the truth — there is no appropriate explanation of the correlation between belief and truth. And, they say, that's a problem – it's a reason to be suspicious of our moral beliefs or to reject non-naturalist realism. Modalists agree that there is no explanatory connection, but claim that this isn't a problem since our beliefs have the appropriate modal features with respect to the truth.

This standoff is central to the modern debate about debunking. Let me put my cards on the table: I don't think third-factor responses to debunking arguments work. (Though, as I noted, that won't be a focus of this paper. I argue against third-factor responses in Bhogal (2022), which is a companion paper to this one.). So, I think, if the explanationist is right there is a good debunking argument against non-naturalist views of morality. And if the modalist is right there is not.

I'm going to argue that the explanationist is right. Or, more carefully, there is a version of the debunking argument that the explanationist is right about, and consequently the explanationist can give a powerful argument against non-naturalist realism. Notably, though, this version of the ar-

[2]There are a couple of other possible responses that I'm putting aside here. Firstly, what Korman (2019, section 5.1) calls *unstable minimalist responses*. Such responses accept that the correlation between moral belief and truth is a lucky coincidence but claim there is good reason to believe that we were, in fact, lucky. This type of response doesn't address the form of debunking argument that I discuss in section 2 and onwards but I won't discuss this further.

Secondly, a more recent response stems from doubts about the distinction between causally inert and causally efficacious entities. (Thanks to a reviewer for pointing me towards this.) A recent talk by Barbara Montero and Clarke-Doane (2022, chapter 2) make the case that mathematical objects are causally efficacious – there is no more problem in them being causal than some of the strange entities postulated by fundamental physics. It's an interesting question how much this thought applies to a non-naturalist conception of morality and whether this suggests that our moral beliefs are non-coincidentally true. But I won't investigate this further here – my focus is on the explanationism vs modalism distinction.

gument differs in some important ways from versions that have received the most attention in the recent literature.

In particular, I argue, if we see debunking arguments as an instance of a broader class of arguments about how we choose between theories and, in the process, deemphasize issues of whether our particular moral beliefs are justified then this allows us to break the standoff between explanationism and modalism.

Understanding the debunking argument in this way allows us to consider cases from science which are analogous to the moral case, seeing whether it is explanatory or modal considerations that are the driving forces behind theory choice in such cases. This, I'm going to claim, favors the explanationist.

1 The Standoff between Modalism and Explanationism

It would be useful to start, though, by clarifying the standoff between modalism and explanationism.

Again, the intuitive thought driving most debunking arguments is that given a non-naturalist conception of morality – where moral facts are mind-independent, non-natural, and causally inert – there is no explanatory 'connection' between our moral beliefs and the truth. And the recognition that our beliefs are disconnected from the truth seems to undermine their justification.[3,4]

I'm understanding modalism as a strategy for defending non-naturalist realism, in light of such arguments, by appealing to facts about the modal features of our moral beliefs. As we noted, the modalist accepts that there is no explanation of the correlation between moral belief and truth, but stresses that our beliefs are *safe* and *sensitive*.[5]

[3] Some debunking arguments focus on knowledge rather than justification – primarily aiming to show that our beliefs do not constitute knowledge, though perhaps knowledge that our beliefs do not constitute knowledge undermines their justification. (Thanks to a reviewer for discussion here.) For now, I'll stick with the idea that debunking aims to undermine justification. Ultimately this won't matter, since my preferred form of argument works very differently.

[4] For simplicity, in the rest of the paper, I'll focus on beliefs about necessary moral facts – like *killing people for fun is wrong* or *an agent in situation X would have reason to* ϕ – and not contingent facts like *Robin shouldn't have shouted at Rohan last week* which are, in some sense, about both the moral and the natural domains.

[5] A related view claims that there is an explanation of the correlation, just an extremely minimal one (see Pust (2005) and Baras (2017)). For example, one could argue that merely giving a causal explanation of why we have the moral beliefs

> **Sensitivity**: Our belief that P is sensitive iff had it been that ¬P, we would not still have believed that P, had we used the method that we actually used to determine whether P.
>
> **Safety**: Our belief that P is safe iff we could not have easily had a false belief as to whether or not Q, where Q is any proposition similar enough to P, using the method that we actually used to determine whether or not P. (Clarke-Doane and Baras, 2019, p. 163)

Consider, for example, my belief that killing people for fun is wrong. Could I have easily had a false belief about this (or about some similar proposition)? It seems not. Such moral truths are necessary – killing people for fun is wrong in all possible worlds – and I could not easily have believed that it is not wrong — I would have to have been an extremely different person to believe that it's acceptable to kill people for fun. So my belief is safe (Clarke-Doane, 2015, section 4.5).

Similarly, my belief seems to be sensitive. It's sensitive if, had it been the case that killing people for fun was not wrong, I would have not believed that killing people for fun is wrong (had I used the method that I actually used to determine whether killing people for fun is wrong). This counterfactual is normally taken to be trivially true, since it could not have been the case that killing people for fun was not wrong — it is metaphysically impossible. (Lewis, 1986, section 2.4).

In much of Clarke-Doane's work (e.g. (2015, pp. 88-92), (2016a, pp. 26-28), (2020, section 5.7)) this argument about sensitivity is only the first horn of a dilemma, posed to debunkers. This horn arises if we assume a 'standard' semantics for counterfactuals. Here's the other horn: If we allow a non-standard semantics, where certain metaphysically impossible worlds – like those where it's morally acceptable to kill people for fun – are relevant to the evaluation of counterfactuals, then it's plausible that the counterfactuals needed for sensitivity will come out false. That is, if killing for fun were not wrong, I would still have believed that it was is wrong. So, it looks like my belief

that we have, and noting that the content of those beliefs is necessarily true, counts as an explanation of the correlation. This type of view won't be the main focus but, as we will see in section 3.1, the core argument of this paper gives reasons to be doubtful of it. (Some might take the question of which views fit into this category and which ones count as 'modalism', as I'm understanding it, to be somewhat fluid, to the extent that the distinction between an 'extremely minimal' explanation, and no explanation at all might be fluid. Thanks to a reviewer for discussion.)

is insensitive. But, Clarke-Doane argues, on this understanding of counterfactuals the demand for beliefs to be sensitive is unreasonable.

For example, if nihilism about composite objects were true, I would still believe that it was false. Consequently, I would still have believed that there was a table in front of me, even when there are just atoms arranged table-wise. So it seems like my beliefs about composition are insensitive. Plausibly, Clarke-Doane argues, this will end up undermining our ordinary beliefs about things like tables, leading to a kind of skepticism.

So, given such a non-standard semantics for counterfactuals, there's a concern that our belief that killing people for fun is wrong will be undermined, since it is insensitive. But, in such a case, we should reject the requirement that our beliefs should be sensitive as being too strong and leading to skepticism.

This is an extremely interesting strategy but, as we will see on page 13, the suggestion that beliefs do not need to be sensitive ends up making things easier for the argument I want to develop. So, I will assume, in the modalist's favor, that our moral beliefs are sensitive, in line with the first horn of the dilemma.[6]

Given that debunkers haven't given us reason to doubt the safety or sensitivity of our beliefs then it's hard to see how debunking arguments could work. When our beliefs are sensitive and safe it's natural to say that they are *reliable*. And once we have reliable beliefs what more could we want? In particular, if we know that our beliefs are safe and sensitive then it seems we are justified in having them. This type of idea – that for our beliefs to be undermined we need reason to think that they are not safe or sensitive – is called *Modal Security* in much of Clarke-Doane's work (e.g. Clarke-Doane and Baras, 2019).

Though, as Clarke-Doane and Baras note, many people find this line of argument extremely unintuitive since it 'implies that one's beliefs can be secure from undermining, even upon learning that they bear no connection to the truth' (p.180). In particular, explanationists, like Korman and Locke

[6]Thanks to a reviewer for discussion.

(2020), Lutz (2020, 2018) and Faraci (2019), think that 'if one learns that there is no "explanatory" connection between the belief and the truth, then that in itself is undermining — regardless of the modal security of the belief' (Clarke-Doane and Baras, 2019, p. 180).

Here's one way to see the disagreement: In ordinary cases of belief we have both the relevant modal conditions and an explanatory connection between belief and truth. Consider my belief that there is a computer in front of me, formed on the basis of my visual perception. It's safe and sensitive, and there is an explanatory connection between my belief and the truth – my belief is explained, in part, by the facts about the computer.

The problem is that in the moral case these modal and explanatory conditions seem to come apart – our beliefs seem to be safe and sensitive without an explanatory connection. Modalists say that since the relevant modal conditions hold we are justified in our moral beliefs. Explanationists say that since there is no explanatory connection between belief and truth we are not justified.[7]

In fact, Clarke-Doane and Baras (section 10) say that the explanationist is misled by the fact that in ordinary cases the modal and the explanatory conditions go together. They say that an explanatory connection is predictive of safety and sensitivity in many cases, but it's really safety and sensitivity that is epistemically valuable. The explanationist, they say, confuses a proxy for what's valuable with the valuable thing itself. But of course, the explanationist could accuse the modalist of a similar mistake – claiming that what's really valuable is genuine explanatory connection between belief and truth and that safety and sensitivity is just a proxy. We seem to have a standoff.

2 Debunking and Theory Choice

The way to make progress on the question of modalism versus explanationism is, I think, to move away from the type of debunking argument we have just been considering – arguments where the

[7] Faraci (2019, pp. 12-13) discusses another case where the modal and explanatory conditions come apart that he claims favors the explanationist. But Clarke-Doane and Baras (2019, p. 175) worry that the case is incoherent. I share this worry.

conclusion is that we lack justification in particular moral beliefs, given non-naturalist realism. Perhaps there is stalemate between the modalist and the explanationist with respect to this argument. But if we consider a slightly different form of debunking argument then a range of new considerations open up.

In particular, it's fruitful, I think, to construe debunking arguments as an instance of more general considerations about theory choice. Doing so will allow us to deemphasize questions about, for example, the precise conditions that determine when our particular moral beliefs – like that killing people for fun is wrong – are justified. Consider, for example, this argument:

(1) Given non-naturalist realism the correlation between our actual moral beliefs and the moral truths is a big coincidence.
(2) There is reason to reject theories that lead to big coincidences.
So,
(3) There is reason to reject either the correlation between our moral beliefs and the truth, or non-naturalist realism.[8]

This argument is very much in the spirit of Field's (1989) argument against mathematical realism.

This type of debunking argument is an instance of a very general consideration — when we have a striking correlation and some theory which implies that the striking correlation is a coincidence, then that's reason to reject the theory or to reject the correlation. In the moral case under investigation the striking correlation is between our beliefs and the truth. But, on this way of understanding the issue, that's rather incidental. The correlation could be between all sorts of things and an argument of this form would still apply.

Imagine a friend told me that she tossed a coin 50 times and it landed heads every time. The theory that it's a fair coin that she tossed normally implies that it's just a big coincidence that the coin landed heads every time. This gives us reason either to reject the theory — to think that it's not a fair coin — or to reject the correlation — to think that my friend was lying and the coin didn't land heads

[8] I discuss how to defend this argument from from third-factor responses in Bhogal (2022).

every time.

We will talk about other cases soon.[9] But, to be clear, the point is not that there can never be any coincidences, or that when a theory implies that a correlation is a coincidence that's a conclusive reason to reject the theory. It's just that coincidences that are sufficiently big can give us reasons to reject a theory, and sometimes those reasons can be rather powerful.

I'll say more about the form of this argument very soon. But first, a natural question arises: What is a coincidence? There are two broad options. Coincidence could be an explanatory notion – some pattern or correlation that has a deficient explanation – or it could be a modal notion – some pattern or correlation that could easily have failed to hold, or had some other modal deficiency. Unsurprisingly, modalists sometimes claim that coincidence is a modal notion and so 'where there is no contingency, there are no coincidences' (Wielenberg (2010, p. 461), see also Clarke-Doane (2020, section 4.5)). Explanationists, on the other hand, typically have explanatory conceptions of coincidence (e.g. Faraci, 2019).

In fact, one way to understand the dispute between modalists and explanationists is as over the nature of coincidence. (Understood in this way, both sides would accept premise (2) of the above argument, but modalists would reject premise (1)).

Ultimately, though, the term 'coincidence' doesn't really matter for the dispute between the modalist and the explanationist – what matters rather is whether modal or explanatory considerations should affect theory choice in such cases. Specifically, what matters is whether modal or explanatory considerations can undermine the conjunction of non-naturalism and the correlation between moral belief and truth. So, I'm going to deemphasize this disagreement over the nature of coincidence. I'll just stipulate that 'coincidence' as I'm using it, is an explanatory notion, and so the modalist strategy, as I'm understanding it – will accept that the correlation between moral belief and truth is a coincidence. They will dispute the argument elsewhere.[10]

[9] But see, for example, Bhogal (2020) and Baras (2022) for many other examples of this type of reasoning – including in important parts of scientific practice.

[10] Thanks to a reviewer for discussion.

This stipulation lines up with most of the modern literature on coincidences. In that literature, it's normally suggested that a coincidence is, roughly, a striking correlation or matching between facts that does not have an explanation (see, for example, Hart and Honoré (1985, p. 74), Lando (2017), Bhogal (2020) and the views that Lando (section 2) calls 'traditional views').

This is only rough. To fill this out into a complete account we would need a story about what makes a correlation striking (see, for example, Horwich (1982, chapter 7), Schlesinger (1987), White (2005), Wong and Yudell (2015), Baras (2022)) and a story about precisely what it takes for a correlation or matching between facts to not have an explanation (Lando, 2017; Bhogal, 2020). These details won't matter for our purposes – we can proceed with this rough characterization of coincidence for now.

A further clarification, about the form of the argument (1)-(3), before we move on. The conclusion is that there is reason to reject either the correlation between our moral belief and the truth, or non-naturalist realism. First and foremost, then, the argument casts doubt on a certain conjunction. There is then a decision about which conjunct to reject. But someone who is committed to non-naturalism is, the argument says, pushed to reject the correlation between moral belief and truth – and therefore to believe that most of our moral beliefs are false. This is clearly a very unpleasant consequence.

Notice, furthermore, that rejecting the correlation between moral belief and truth is not the same as believing that particular moral beliefs – like killing people for fun is wrong – are unjustified or don't constitute knowledge. My conclusion is that the non-naturalist should reject the correlation between moral belief and truth – and so should think that many of our moral beliefs are false.

But I want to be agnostic regarding further claims about the justification of particular moral beliefs. Perhaps it is possible to be justified in believing that killing people for fun is wrong (and other similar moral beliefs) while believing that most of our moral beliefs are false. Whether this is possible comes down to questions about the force of higher-order evidence. And perhaps it is even possible to be justified in believing that killing people for fun is wrong while believing that there is no explana-

tory connection between belief and truth, just as the modalist claims. My argument doesn't make commitments one way or the other here.

Rather, my conclusion is simply that the non-naturalist is pushed to reject the correlation between moral belief and truth – and therefore to believe that most of our moral beliefs are false. This looks like a major challenge to non-naturalism.

2.1 The Modalist response to (1)-(3)

So, the argument (1)-(3) has a different form from many of the debunking arguments that have received the most attention in the recent literature[11] – it's not about whether non-naturalism is consistent with the justification of particular moral beliefs. But this shift doesn't, of course, mean that the modalist has to give up. A modalist can respond to (1)-(3) in a way that is closely analogous to their response to other versions of the argument. In this section we will consider how that response will go.

Notice that the argument (1)-(3) embodies an explanationist position. It says that we have reason to reject the combination of the correlation between our moral beliefs and the truth and non-naturalist realism because that combination would make the correlation a coincidence. And the correlation being a coincidence is an *explanatory* deficiency — it's that the striking correlation is unexplained.

The modalist move we discussed in the last section is to argue that even though moral belief and truth seem to be explanatorily disconnected — that is, the correlation between them is unexplained — this isn't epistemically important. Rather, what is important is that our moral beliefs are safe and sensitive.

How can the modalist make this kind of move with respect to (1)-(3)? Here, the explanationist claims that we have reason to reject the correlation between belief and truth or non-naturalist realism because the correlation is not explained. The modalist move – analogous to their responses to other debunking arguments – will be to accept that the correlation is not explained but to claim that

[11] Though certainly this type of argument has received attention, e.g. Field (1989), Enoch (2010), Schechter (2010).

this doesn't provide a reason to reject the correlation or non-naturalist realism because, given non-naturalist realism, the correlation between belief and truth has certain appropriate modal features. This move is, in effect, a rejection of premise (2).

2.2 Appropriate Modal Features

There is, though, an obvious question: Just what are these appropriate modal features?

The standard modalist move is that the safety and sensitivity of our beliefs make them unproblematic. But, as we noted in the last section, our version of the debunking argument is not distinctively about belief.[12] Rather, it's an instance of a more general argument that when we have a striking correlation and some theory which implies that the striking correlation is a coincidence, then that's reason to reject the theory or to reject the correlation. This argument applies to correlations that have nothing to do with belief.

The issue is that safety and sensitivity *are* distinctively about belief. So what the modalist needs are generalizations or analogues of safety and sensitivity which apply not just to the correlation between our beliefs and the truth, but to correlations or matchings between facts more generally. Roughly speaking, the idea of safety is that our beliefs could not easily have been false. And the idea of sensitivity is that if the facts had been different our beliefs would have been different.

It's easy enough to see how to extend these ideas to correlations more generally. Take, for example, a correlation or matching between fact A and fact B.

Safety*: The correlation is safe* if and only if it could not easily have failed to hold, that is, if A could not easily have failed to match B.

Sensitivity*: The correlation is sensitive* if and only if had it been the case that one side of the correlation did not hold the other would not have held. That is, if A did not hold then B would not have held and vice versa.

Safety* and Sensitivity* are the natural candidates for the features that the modalist will rely on.

[12] Though, of course, the argument does have implications for what we should believe.

3 Protons and Electrons

Perhaps it may seem like this shift to debunking arguments like (1)-(3) hasn't helped us at all. There is still, it seems, a standoff between the explanationist and the modalist. The explanationist says that the correlation between moral belief and truth is unexplained given non-naturalist realism, and that's a problem with either the correlation or non-naturalist realism. The modalist replies that it's not a problem because the correlation has the appropriate modal features — in particular safety* and sensitivity*.

But moving to arguments like (1)-(3) in fact opens up a range of new cases that we can use to evaluate explanationism and modalism. In particular, since (1)-(3) is an instance of general considerations about theory choice, which are not distinctively about belief, then the cases we can use to evaluate (1)-(3) don't have to be correlations between belief and truth. In this section I'm going to appeal to a scientific case that, I will argue, ends up favoring the explanationist.

The case involves a striking correlation and a theory which implies that the correlation is safe* and sensitive*. So, the modalist should say that the combination of the theory and the correlation is unproblematic. Nevertheless, it seems clear that we should reject either the correlation or the theory because the correlation would be unexplained given the theory.

I'll focus on a simple case, but later I'll point toward how it's illustrative of more complicated cases which are currently under dispute in scientific practice. Here is the case:

Protons and Electrons Protons are positively charged. Electrons are negatively charged. However, the absolute value of their charge is the same — call this fact the *charge correlation*. Specifically, protons have a charge of $1.602176634{\times}10^{-19}$ coulombs, while electrons have a charge of $-1.602176634{\times}10^{-19}$ coulombs.

The charge correlation is very striking. Consider a theory, T, that implies that it is just a complete coincidence that the charges are the same – that there is no explanation of the matching. This seems like a big problem for T. Perhaps it wouldn't be a conclusive reason to reject T – if T is successful

enough we might still be justified in accepting it – but it's a bad-making feature of the theory.

Now let's imagine what such a theory, T, might look like. Let T state that it is a basic law of nature that protons have a charge of $1.602176634{\times}10^{-19}$ coulombs, and also it is a basic law of nature that electrons have a charge of $-1.602176634{\times}10^{-19}$ coulombs. And T has nothing more to say about why the particles have these charges.[13]

Importantly, given T the charge correlation is nomically necessary.

Such a case, I claim, provides an argument against the modalist. Given T there is no explanation of the charge correlation but the correlation seems to be safe* and sensitive*. We still have reason, however, to reject either the charge correlation or T. So, it is the explanatory factors — the way in which, given T, the charge of the proton and electron are explanatorily disconnected — that are relevant for theory choice, and not the modal features of the correlation. It is the explanationist who is right about arguments of the form (1)-(3), not the modalist.

Notice that it's now clear why the modalist strategy of accepting that our moral beliefs are insensitive but denying that this is a problem, discussed on page 5, doesn't help with respect to this argument. This is because our argument doesn't work by demanding that beliefs must be sensitive. Rather, the argument, here, is that the modalist can't claim that the correlation between moral belief and moral truth is unproblematic because it has some appropriate modal features – notably safety* and sensitivity* – since the charge correlation has these features too. Any moves that weaken what these appropriate modal features are, e.g. by saying that sensitivity and sensitivity* don't matter, make it easier to argue that the charge correlation has these features. So I'll continue to assume, in the modalist's favor, that our moral beliefs are sensitive.

The best way to explain and defend this argument is, I think, by considering some possible modalist

[13]This won't matter for what's to come, but our modern physics is not like T – it gives us an outline of an explanation for why protons and electrons have the same charge. Or, at least, it suggests that the charge correlation is somewhat less surprising and striking. It's to do with the fact that charges are quantized – charge seems to be a discrete rather than a continuous quantity. This makes it much less surprising that elementary particles, or particles that consist in simple combinations of elementary particles, like protons and electrons, match in their charge.

There is a further question about exactly why charges are quantized. This isn't clear, but the most accepted story is based on Dirac quantization. The physics here is unsettled but situation looks very different from what is postulated by theory T.

responses. I'll do that in the rest of the section. So how might a modalist respond? Firstly, they could claim that there is an explanation of the charge correlation. Secondly, they could claim that the correlation is not safe* or sensitive*. Thirdly, they could claim that we do not, in fact, have reason to reject the combination of the correlation and T. Let's consider these possible responses in turn.

3.1 No Explanation

Let's start with the claim that given theory T there is no explanation of the correlation between the charges. This, I take it, is fairly intuitive — after all, T doesn't seem to provide any connection between the charge of the proton and the electron. But there is a natural way that someone might deny it. They might claim that charge correlation is explained as follows: The fact that it's a basic law that electrons have a charge of $-1.602176634{\times}10^{-19}$ coulombs explains why electrons have a charge of $-1.602176634{\times}10^{-19}$ coulombs. The fact the it's a basic law that protons have a charge of $1.602176634{\times}10^{-19}$ coulombs explains why protons have a charge of $1.602176634{\times}10^{-19}$ coulombs. $-1.602176634{\times}10^{-19}$ and $1.602176634{\times}10^{-19}$ have the same absolute value. All this together explains why the absolute value of the charges of protons and electrons match.

I'll make three points about this strategy.

Firstly, claiming that there is an explanation of the charge correlation isn't really a way for the modalist to defend their position. The result that we want is that we should reject theory T. But, of course, claiming that theory T does in fact explain the charge correlation doesn't help us get this result. What this is, rather, is a criticism of the explanationist, saying that they face problems with the protons and electrons case too.[14]

Secondly, the focus of this paper is on the modalist strategy of accepting that there is no explanatory connection between our moral beliefs and the moral truths but claiming that this doesn't matter,

[14]This reasoning also allows us to see a major problem with the view mentioned in footnote 5. The view was that extremely 'minimal' explanations of the correlation between moral belief and truth can allow the non-naturalist to happily accept the correlation. But, the existence of a similar minimal explanation of the charge correlation should not make us happy to accept both the charge correlation and theory T.

since our beliefs have the relevant modal features with respect to truth. Consequently, it's fair for us to assume that the modalist will similarly take there to be no explanation with respect to the charge correlation.

Thirdly, it's intuitively clear that this proposed explanation of the charge correlation is deeply unsatisfying and explanatorily deficient. It doesn't really explain the matching between the charges, rather it merely separately explains the precise value of the charge of the proton and the precise value of the charge of the neutron. Similarly, separately explaining why each coin tossed landed heads, by giving the microphysical details of exactly how each coin was tossed, is not a satisfying explanation of why the coin landed heads every time my friend tossed it.[15] Consequently, the explanationist should be happy to accept that this proposed 'explanation' doesn't count as an explanation of the charge correlation. (Or, alternatively, they could accept that it does, strictly speaking, count as an explanation, but even in light of this type of explanation the charge correlation is still a huge coincidence.) Making precise the exact sense in which the proposed explanation is deficient is an interesting and complicated task.[16] Going into it further would take us too deep into the literature on scientific explanation. But still, it's easy to recognize that the proposed explanation is deficient.

For these reasons, claiming that the charge correlation has an explanation, given T, is not a promising way for the modalist to defend themselves.

3.2 Modal Features

The next possible response: Given T, is the correlation between the charge of the proton and electron safe* and sensitive*?

[15] A reviewer suggests that perhaps the fact that the correlation in the moral case involves contingent moral beliefs, unlike the charge correlation which is between two nomically necessary facts, means that the analogous explanation – separately explaining the moral facts and our moral beliefs – is more satisfying. This is an interesting suggestion, but I don't share the intuition that such contingency helps – this coin case is a contingent correlation, but still separately explaining the tosses seems deeply unsatisfying. And this judgement is shared by others in the philosophy of science literature on explanation (e.g. Owens (1989, chapter 1), Lando (2017, pp. 144-5), Bhogal (2020, section 4)).

[16] See Field (1996, section V), Lange (2010), Tersman (2016, section 3), Baras (2017), Faraci (2019), Bhogal (2020) and Baras (2022, chapter 4) for discussion.

3.2.1 Safety*

It's pretty clear that the correlation is safe*. T implies that it is nomically necessary that the absolute value of the charge of the electron matches that of the proton — the correlation is guaranteed by the laws of nature. In this clear sense, then, it could not easily have failed to hold because that would require the laws to have been different.

The view that the laws of nature are modally robust and so could not easily have been false is, I take it, the standard position. For example, it's standardly taken to be the case that a possible world that contains widespread violations of the actual laws of nature is, in virtue of that, very distant and dissimilar from the actual world (Lewis, 1979). And this distance from the actual world makes it the case that it could not easily have held.

Furthermore, we have reason to think that the modalist, in particular, would accept that the laws could not easily have been false. As we discussed in section 1 the modalist typically argues that my beliefs about killing people for fun being wrong could not easily have been false. The reasoning is that the closest world where I have a false belief about killing people for fun being wrong is one where I believe that it is acceptable. But I could not easily have had such a belief — I would have to have been an extremely different person to do so.

But if I could not have easily believed that it's ok to kill people for fun — something that would have occurred if my upbringing had been very different, or if I was hypnotized, or if I was in the grip of a deeply mistaken philosophical theory — then it seems clear that it could not easily have been the case that the basic physical laws are different. And so, given T, it could not easily have been the case that the correlation between the protons and electrons failed to hold.

3.2.2 Sensitivity*

What about sensitivity*? Can the modalist respond by denying that the charge correlation is sensitive*?

Whether the charge correlation is sensitive* is a harder question than whether it is safe*. It's particularly hard because the counterfactuals we have to evaluate are *counternomics* — that is, they have nomically impossible antecedents. They ask what would happen if electrons or protons had a different charge. Often it's extremely hard for us to evaluate what would be the case if the laws were different. What would be the case, for example, if the world were not relativistic? Would quantum mechanics be true? Would the world be Newtonian? It's hard to see what considerations we could bring to bear upon this.[17]

It's hard to know, in particular, in a situation where the charge of electrons is nomically necessary, what would be the case if the charge was different. So, does that mean that the modalist can reasonably claim that the correlation fails to be sensitive* and, further, that this the reason that we are inclined to reject the combination of the correlation and T?

I don't think this is an attractive option. The reason is that there are background assumptions we can add which guarantee that sensitivity* holds, but these background assumptions don't make a difference to how we should react to the conjunction of the theory T and the charge correlation — we should still reject that conjunction.

In particular, it's very plausible that electron and protons have their charge essentially and, therefore, as a matter of metaphysical necessity.

Assuming this, for a moment, consider the counterfactuals relevant for the sensitivity of the charge correlation, for example: If the charge of the proton hadn't been $1.602176634{\times}10^{-19}$ coulombs then the charge of the electron would not have been $-1.602176634{\times}10^{-19}$ coulombs.

As we discussed in section 1 the modalist takes the counterfactual *if it had been the case that killing people for fun was not wrong then I would have not believed that it is wrong* to be trivially true, since it could not have been the case that killing people for fun is not wrong. (A reminder about the dialectic here: As we discussed in section 1, the modalist doesn't have to commit to the trivial truth of this counterfactual (and other counterpossibles). But if they don't then it looks like they have to deny

[17] This is not to say that every case is like this. There are some counternomics which are easier to evaluate and which seem clearly true (Tan, 2019).

that sensitivity is requirement for justified belief because this counterfactual will end up false. And as we discussed on page 13 this move makes my argument easier to develop. So we are, in the modalist's favor, assuming triviality at this point.)

Similarly, if the charge of protons is necessary then the counterfactual *if the charge of the proton hadn't been* $1.602176634{\times}10^{-19}$ *coulombs then the charge of the electron would not have been* $-1.602176634{\times}10^{-19}$ *coulombs* would be trivially true.

The same is true for the counterfactual *if the charge of the electron hadn't been* $-1.602176634{\times}10^{-19}$ *coulombs then the charge of the proton would not have been* $1.602176634{\times}10^{-19}$ *coulombs*. The modalist's prior reasoning implies that it is trivially true. So, if the charges of protons and electrons are essential, and therefore metaphysically necessary, then the charge correlation is sensitive*, given T.

But why think that protons and electrons have their charge essentially? One reason is if we accept a nomic essentialist position which implies that the laws of nature are metaphysically necessary. There are a variety of different approaches to laws that would get that result. The most common version of the view appeals to the nature of the dispositions or powers that properties have essentially (e.g. Shoemaker (1980); Ellis (2007); Bird (2007)); other versions appeal to claims about how possible worlds are generated (e.g. Wilson (2020)).

But even if we don't hold any of these positions, it's still natural to think that electrons and protons have their charge necessarily for broadly Kripkean reasons. If there was some possible world where there was something electron-like but it was, in fact, positively charged rather than negatively charged then we would take it to be a different particle, not an electron.[18]

Getting into the detail of these positions that imply that protons and electrons have their charge essentially would take us too far afield. But importantly, these positions imply that the charge correlation is sensitive*.

Can't the modalist simply deny these positions though? Can't they deny these claims about scientific

[18] Of course, certain discoveries about the actual world could lead us to doubt such a claim – Kripkean necessities are always fixed by actual world facts – but holding fixed our current understanding of electrons this claim is plausible.

essentialism and the Kripkean considerations, saying that the charge correlation isn't sensitive* given T and that's the reason we should reject the theory T?

They could, but the resulting view is not attractive. It makes the acceptability of T, given the charge correlation, depend upon background philosophical views about the nature of scientific laws, or about what properties are essential. But this isn't right. Whether we should accept the combination of the charge correlation and theory T is a first order scientific question – one that should not be affected by the truth of the philosophical theories about laws and modality. In particular, we should continue to reject T, given the charge correlation, regardless of whether the charges are had essentially.

So the strategy of denying sensitivity* is not, I think, an attractive way for the modalist to resist the argument.

3.3 Don't Reject T

Another possible modalist response is that we should be happy to accept the conjunction of the charge correlation and T. This, they could claim, is not a bad result.

This, I think, is implausible. That the charge of the proton and the charge of the electron seem unrelated given T is a powerful reason to think that T is unlikely. But we can go further by noting that the structure of **Protons and Electrons** is the same as other cases that are under serious scientific investigation. Those cases suggest that we should reject T.

In particular, the 'hierarchy problem' of the mass of the Higgs Boson is a case of this kind. Going into detail about this problem would require a long discussion of particle physics and, of course, now is not the time for that. But the problem is, again, one which involves a strikingly coincidental match between two values, both of which plausibly hold with nomic necessity. Very roughly speaking, the matching is between the different factors that contribute to the mass of the Higgs Boson. The bare mass of the Higgs Boson and the radiative corrections appear to match, and therefore to cancel, in a seemingly miraculous way leaving the total mass extremely small (see, e.g. Friederich (2022),Williams (2015), Torrente-Lujan (2014), Craig (2020), Hossenfelder (2021)).

Importantly, the hierarchy problem is taken to be a *problem*. The matching between these values is a problem for our current theories and a reason to look for new physics. Nomic necessity of the matching isn't a reason to be satisfied.

The response that we shouldn't reject T goes against how similar cases are dealt with in scientific practice. Consequently, we shouldn't accept this response.

3.3.1 What this case tells us

The takeaway from cases like **Protons and Electrons** is that explanatory factors are most relevant to theory choice, not modal ones. To be clear, we don't need to say that modal considerations are *never* relevant to theory choice. The point is that even when a correlation satisfies the analogues of safety and sensitivity, we still have reason to reject either the correlation or the theory that implies that the correlation has no explanation.

The modalist response to (1)-(3), then, is unsuccessful. The explanatory disconnection between moral belief and truth is a reason to reject the combination of the correlation between belief and truth and non-naturalist realism — this is not undermined by pointing out that the correlation has certain modal features. It is the explanationist that wins out over the modalist.

Of course, philosophical arguments like this are rarely conclusive – there are still further avenues for resistance. I'll finish by mentioning two such avenues (of course, this is in addition to the variety of responses that I considered in section 3).

Firstly, the modalist could fight back by appealing to scientific practice – claiming that the role of explanation in scientific theory choice isn't as significant as I've suggested and perhaps modal considerations are more significant. This is a difficult task – I suspect that it's unlikely that modalist would succeed. (See, for example, Emery's (2022) discussion of the 'pattern-explanation principle' for more cases that support the importance in scientific practice of explaining striking correlations.) But, nevertheless, I really welcome efforts of this kind. Seeing that this kind of defense of non-naturalist realism pushes us toward certain commitments about explanation in scientific practice is,

I think, an important step in the literature.[19]

Secondly, perhaps the modalist could respond by claiming that it's not safety* and sensitivity* that matters for whether we should accept a correlation, rather it's some other modal features. They could claim, that is, that there are some modal features had by the correlation between moral beliefs and truths but are not had by the correlation in **Protons and Electrons** and analogous scientific cases. And, further, that those particular modal features are importantly connected to theory choice — they are the reason that we are pushed to reject the theory T in **Protons and Electrons** but should be happy to accept non-naturalist realism about morality. This is certainly a pathway for the modalist, but it seems to be a very difficult task to find such modal features. Perhaps, in line with a suggestion a reviewer makes, the best option for the modalist is to try to exploit the fact that the charge correlation is between two (nomically) necessary facts, while in the moral case, the correlation is between necessary moral facts and contingent moral beliefs. It might be possible to leverage this disanalogy into a view of what such modal features would be. But, at least for me, it's very hard to see how this would work. It's a challenge to the modalist to be specific about what such features are.

4 Conclusion

The modalist response to (1)-(3), I have argued, fails. We should reject the conjunction of non-naturalist realism and the correlation between moral belief and truth. This is a major problem for the committed non-naturalist.

As we noted, this is importantly different from many prior arguments which conclude that we should give up non-naturalism or our particular moral beliefs – like that killing people for fun is wrong. My argument is silent about such beliefs.

The strategy was to consider the methodology in scientific cases like **Protons and Electrons** and the hierarchy problem – leveraging that into a rejection of the conjunction of non-naturalism and the correlation between moral belief and truth.

[19] Thanks to a reviewer for discussion here.

But this strategy invites a further question: Just what are the underlying epistemic principles that explain *why* we should reject that conjunction? I give reasons to think *that* we should reject that conjunction but not a full explanation of *why*. Of course, a large part of the story is that there is a coincidence-avoidance norm. But just how does a coincidence-avoidance norm fit into the larger epistemic picture? More specifically, does an application of this norm give us evidence that rebuts any initial evidence we have for the conjunction of non-naturalism and the correlation between moral belief and truth? Or is that initial evidence undermined? And if it is undermined, do we have reason to think that a belief in the conjunction is not safe and sensitive?

These are great questions (and I'm taking some of them on in in-progress work) but they are not ones that I can decide here. And it's not necessary to decide them, given the project of this paper. We can know that we should reject either non-naturalism or the correlation between moral belief and truth without knowing the details of the underlying epistemology. In fact, this is a big advantage of the strategy that the paper implements: if we trust the methodology of science we can use it to resolve disputes – like that between explanationism and modalism – without having to derive the result from an underlying epistemic theory.

References

Baras, D. 2017. Our reliability is in principle explainable. *Episteme*, 14(2):197–211.

Baras, D. 2022. *Calling for Explanation*. Oxford University Press.

Bhogal, H. 2020. Coincidences and the Grain of Explanation. *Philosophy and phenomenological research*, 100(3):677–694.

Bhogal, H. 2022. What's the coincidence in debunking? *Philosophy and phenomenological research*.

Bird, A. 2007. Nature's Metaphysics. *OUP*.

Clarke-Doane, J. 2015. Justification and explanation in mathematics and morality. *Oxford studies in metaethics*, 10:80–103.

Clarke-Doane, J. 2016a. Debunking and Dispensability. In Leibowitz, U. D. and Sinclair, N. (eds), *Explanation in Ethics and Mathematics: Debunking and Dispensability*. Oxford University Press.

Clarke-Doane, J. 2016b. What is the Benacerraf problem? In *Logic, Epistemology, and the Unity of Science*, pp. 17–43. Springer International Publishing, Cham.

Clarke-Doane, J. 2020. *Morality and Mathematics*. Oxford University Press.

Clarke-Doane, J. 2022. Mathematics and Metaphilosophy. In *Elements in the Philosophy of Mathematics*. Cambridge University Press.

Clarke-Doane, J. and Baras, D. 2019. Modal Security. *Philosophy and phenomenological research*, 19:1.

Craig, N. 2020. Naturalness Hits a Snag with Higgs. *Physics*, 13.

Ellis, B. 2007. *Scientific Essentialism*. Cambridge University Press.

Emery, N. 2022. The Governing Conception of Laws. *Ergo*.

Enoch, D. 2010. The epistemological challenge to metanormative realism: how best to understand it, and how to cope with it. *Philosophical studies*, 148(3):413–438.

Enoch, D. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. OUP Oxford.

Faraci, D. 2018. Explanation in Ethics and Mathematics: Debunking and Dispensability By Uri D. Leibowitz and Neil Sinclair. *Analysis*, 78(2).

Faraci, D. 2019. Groundwork for an explanationist account of epistemic coincidence. *Philosophers Imprint*.

Field, H. 1989. *Realism, Mathematics & Modality*. Realism, Mathematics & Modality. Basil Blackwell, Oxford.

Field, H. 1996. The a prioricity of logic. In *Proceedings of the Aristotelian Society*, vol. 96, pp. 359–379.

Friederich, S. 2022. Fine-Tuning. In Zalta, E. N. (ed), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Hart, H. L. A. and Honoré, T. 1985. *Causation in the Law*. OUP Oxford.

Horwich, P. 1982. *Probability and Evidence*. Cambridge University Press.

Hossenfelder, S. 2021. Screams for explanation: finetuning and naturalness in the foundations of physics. *Synthese*, 198(16):3727–3745.

Korman, D. Z. 2019. Debunking arguments. *Philosophy Compass*, 14(12):e12638.

Korman, D. Z. and Locke, D. 2020. Against Minimalist Responses to Moral Debunking Arguments. *Oxford Studies in Metaethics*, 15.

Lando, T. 2017. Coincidence and Common Cause. *Noûs*, 51(1):132–151.

Lange, M. 2010. What Are Mathematical Coincidences (and Why Does It Matter)? *Mind*, 119(474):307–340.

Lewis, D. 1979. Counterfactual Dependence and Time's Arrow. *Nous*, 13(4):455–476.

Lewis, D. 1986. *On the Plurality of Worlds*. Basil Blackwell, London.

Lutz, M. 2018. What makes evolution a defeater? *Erkenntnis*, 83(6):1105–1126.

Lutz, M. 2020. The Reliability Challenge in Moral Epistemology. *Oxford Studies in Metaethics*, 15.

Owens, D. 1989. Causes and Coincidences. *Proceedings of the Aristotelian Society*, 90:49–64.

Pust, J. 2005. On explaining knowledge of necessity. *Dialectica*, 58(1):71–87.

Schafer, K. 2017. Review of Leibowtiz and Sinclair (Eds.) Explanation in Ethics and Mathematics: Debunking and Dispensability. *Notre Dame Philosophical Reviews*, 2017.

Schechter, J. 2010. The Reliability Challenge And The Epistemology Of Logic. *Philosophical Perspectives. A Supplement to Nous*, 24(1):437–464.

Schlesinger, G. N. 1987. Miracles and Probabilities. *Noûs*, 21(2):219–232.

Shoemaker, S. 1980. Causality and Properties. In Van Inwagen, P. (ed), *Time and Cause: Essays presented to Richard Taylor*, pp. 109–135. Springer Netherlands, Dordrecht.

Skarsaune, K. O. 2011. Darwin and moral realism: survival of the iffiest. *Philosophical studies*, 152(2):229–243.

Tan, P. 2019. Counterpossible Non-vacuity in Scientific Practice. *The journal of philosophy*, 116(1):32–60.

Tersman, F. 2016. Explaining the reliability of moral beliefs. In Leibowitz, N. S. A. (ed), *Explanation in Ethics and Mathematics: Debunking and Dispensability*. Oxford University Press.

Torrente-Lujan, E. 2014. The Higgs mass coincidence problem: why is the Higgs mass $m_H^2 = m_Z m_t mH2 = mZmt$? *The European Physical Journal C*, 74(2):2744.

White, R. 2005. Explanation as a Guide to Induction. *Philosophers' Imprint*, 5:1–29.

Wielenberg, E. J. 2010. On the Evolutionary Debunking of Morality. *Ethics*, 120(3):441–464.

Williams, P. 2015. Naturalness, the autonomy of scales, and the 125GeV Higgs. *Studies in History and Philosophy of Science. Part B. Studies in History and Philosophy of Modern Physics*, 51:82–96.

Wilson, A. 2020. *The Nature of Contingency: Quantum Physics As Modal Realism*. Oxford University Press.

Wong, W.-H. and Yudell, Z. 2015. A normative account of the need for explanation. *Synthese*, 192(9):2863–2885.