Moral Principle Explanations of Supervenience*

Harjit Bhogal

Here's a very common challenge for defenders of non-naturalism about morality. The moral facts supervene upon the natural facts — there cannot be a change moral facts without a change in the natural facts. More specifically, a certain action with certain moral properties couldn't have had some other moral properties without having different natural properties (including, perhaps extrinsic or relational natural properties). And similarly for other bearers of moral properties, not just actions.

Here's a common formalization of this idea:

$$\Box(\forall F \in M)\forall x(Fx \to (\exists G \in N)(Gx \land \Box\forall y(Gy \to Fy)))$$

where M is the set of moral properties and N is the set of natural properties.

This principle says that: For every moral property F, if something has F then it has a natural property G such that it's metaphysically necessary that anything that is G is F.

Perhaps some nuances need to be added to this formulation (see, for example, Atiq (2020)), but those don't matter for our purposes.

Explaining supervenience is a central challenge for the non-naturalist. The supervenience principle we just wrote down describes a very specific type of moral choreography — the moral and the natural properties dance together in a highly regimented way. But, the non-naturalist says, moral properties are sui generis. They are completely distinct from the natural properties. So, on the face of it, it's not easy to see what explains this choreography.[1]

But there is a non-naturalist strategy which purports to explain supervenience rather simply. So simply, in fact, that Fogal and Risberg (2020, p. 177) say that, in light of it 'the attention garnered by Strong Supervenience (and its ilk) in recent decades looks rather misplaced'. They call it the *Divide and Conquer* strategy, and I will follow them in that terminology.

---

[1]Perhaps explaining supervenience is also a challenge for other views of the metaphysics of morality, but I'll focus on non-naturalism.

The divide and conquer strategy was most influentially defended by Scanlon (2014).[2] The strategy starts with claiming that particular moral facts are explained by general moral principles, along with matters of natural fact. The way that Fogal and Risberg (p. 175) put it is that moral explanations have this structure:

Explanans: particular natural fact(s) (e.g. a is a lie).
Principle: general explanatory moral principle (e.g. lying is wrong).
Explanandum: particular moral fact (e.g. a is wrong).

Given that all particular moral facts are explained in this way there is a quick argument to supervenience. Moral facts are either general explanatory moral principles or particular moral facts. If we assume, as seems reasonable to most, that the general moral principles are necessary, then, trivially, they can't change without the natural facts changing, because they can't change at all.[3] Particular moral facts, on the other hand, are determined by the moral principles and the natural facts, as per the structure of moral explanations just described. And since the moral principles can't change, this implies that the particular moral facts can't change without the natural facts changing. So both the particular and the general moral facts can't change without the natural facts changing. So, the moral facts can't change without the natural facts changing. As long as there are such moral principles, then, supervenience looks extremely unpuzzling.

In Scanlon's development of the strategy the general moral principles are of the form $\forall x (R(p, x, c, a))$, where the holding of this relation means, very roughly, that if proposition p were true and agent x was in circumstance c, then one reason to do action a would be p. These principles are 'pure' moral facts and are necessary. The 'mixed' moral facts are explained by these pure facts and the natural facts.

---

[2] Fogal and Risberg also compare the strategy to Enoch (2011, section 6.2.2). Though this comparison is complicated by the fact that Enoch distinguishes two distinct supervenience theses and gives them separate explanations. *Specific supervenience* 'is the supervenience of the normative properties on the natural ones *that they do in fact supervene on*'. So, for example, if Act Utilitarianism is the true moral theory then specific supervenience is the supervenience of the moral properties of actions on the facts about utility. It is specific supervenience that Enoch explains in a way similar to the the divide and conquer strategy. *General supervenience*, on the other hand — 'the supervenience of the normative properties on *whatever natural properties they supervene on*' — is explained very differently by Enoch. It's general supervenience that is the focus of most of the literature on moral supervenience, and the focus of this paper. So it's not really clear that Enoch is best interpreted as implementing a divide and conquer strategy.

[3] Fine (2002) and Rosen (2020, 2021) reject the idea that such moral principles are metaphysically necessary, instead claiming that they are only *normatively necessary*. If we accept this view then the divide and conquer strategy would not establish the supervenience of the moral on the natural in the space of metaphysically possible worlds, but it would establish supervenience in the space of normatively possible worlds. (Interestingly, Scanlon (2014, p.41) in places, seems to subscribe to this view of the modal status of moral principles but the interpretation isn't particularly clear (see Rosen (2017a, sections VI-VII)).) However, these issues are largely orthogonal to the discussion of the paper so we will simply assume that moral principles are metaphysically necessary. But these issues will come up in passing a couple of times.

This strategy is, I think, extremely compelling and attractive. Just with one piece of metaphysics — explanatory moral principles — we get supervenience very quickly. And the postulation of such moral principles is very intuitive and attractive for the non-naturalist.

Here's one way to see the attractiveness of such moral principles: The moral properties don't seem to be distributed they way they are by accident. Consider many instances of lying, action A, B, C …. It turns out that they are all (pro tanto) wrong. But, it seems, this regularity can't be coincidental — it doesn't just happen to be the case that A, B, C, and so on are wrong. There has to, it seems, be an explanation. The most intuitive explainer to appeal to, for the non-naturalist, is some moral law or principle that lying is wrong.

This reasoning is very similar to one that commonly motivates postulating laws of nature. When we look at many instances of massy objects being attracted to each other, for example, we think that this regularity can't be coincidental — there must be, that is, some law or principle that explains this regularity (see, for example, Armstrong (1983, pp. 51-60), Strawson (1989, pp. 25-29)). (More specifically, this is a motivation for postulating *non-Humean* laws — where the laws are over and above the patterns of events. On a Humean conception of laws, where laws just are patterns of events, it's somewhat hard to see how the laws could explain the pattern. Though the debate here is very rich and complicated (see Bhogal (2020, section 1)) for a survey.)

The exact metaphysical nature of these moral principles is a complicated issue, in the same way that the metaphysical nature of scientific laws is complicated and controversial (see, e.g. Robinson (2011), Elliott (2014), Rosen (2017b), Fogal and Risberg (2020, sections 4-5)). The debate is rich and interesting but such details, broadly speaking, won't matter for our discussion of supervenience.

There are, of course, other strategies for explaining supervenience. But the divide and conquer strategy is particularly interesting because of the way it makes supervenience seem very obvious and to be expected, given metaphysical assumptions that are very simple and intuitive for the non-naturalist. And, consequently, it doesn't seem like the supervenience challenge is something we should focus metaethical discussion on.

Compare this to, for example, Leary's (2017) explanation of supervenience that postulates the existence of properties with distinctive 'hybrid' essences that connect the natural to the moral. This is a very interesting proposal, but it certainly doesn't make supervenience seem unimportant or unworthy of investigation. Rather, supervenience is the key challenge around which the whole metaphysics is built.

So the divide and conquer strategy is a very compelling strategy for explaining perhaps the biggest metaphysical puzzle with non-naturalist realism, and it's based upon a very intuitive appeal to moral principles. However, I'm going to argue that this strategy fails. And, further, investigating this

strategy will tell us a lot about the prospects for a non-naturalist realism based on such metaphysical principles.

In particular, I'm going to argue that there are two ways to interpret the divide and conquer strategy but neither of them are successful. The first strategy doesn't give an adequate explanation of supervenience, while the second perhaps does give an adequate explanation, but only by introducing something similarly puzzling that needs explanation.

1   Two options for explaining supervenience

Let's look at the divide and conquer strategy in more detail. The strategy involves claiming that supervenience follows from the existence of these explanatory moral principles. We haven't, though, been particularly clear on what, precisely, the explanation of supervenience is. In particular, what is the explanans? There are, I think, two possible options.

Firstly, supervenience could be explained by the actual moral principles that hold. For simplicity, in the discussion to follow let's assume that act utilitarianism is true — so the relevant moral principle is that actions are morally right if they maximize utility. Call this principle AU. So the first option is that supervenience is explained by AU. (As we discussed, on Scanlon's version of the strategy the moral principles have the form $\forall x(R(p, x, c, a))$, not the form of AU. But this won't matter for the discussion to come, so it's simpler to assume that AU is the true moral principle.) Secondly, supervenience could be explained not by the actual moral principle AU, but by some generic features of the actual moral principles.

This distinction might seem somewhat puzzling — after all, the features of the actual moral principles are entailed by AU, since AU just is the actual moral principle. But, distinguishing these types of explainers — distinguishing, that is, between explanations that appeal to the full details of a situation and those that appeal only to some general features of the situation — is extremely common and important in the literature on scientific explanation. Consider, for example, the classic case of Putnam's Peg (Putnam, 1975). Imagine I try to put a square peg with 1 inch sides into a round hole with a 1 inch diameter. I, unsurprisingly, fail. One way to explain this is to appeal to the precise details of my attempts to put the peg into the round hole, appealing for example, to every fundamental particle and force in the system, and the fundamental laws of nature, in order to show how my attempts led to failure. But another way to explain this is not to appeal to the full details of the situation, and the full details of the laws of nature, but rather to very generic features of the situation — in particular the geometric features of the peg and the hole and the relevant features of the laws that lead to the solidity of the peg and the hole. Notice that this second explanation, although

it appeals to features that are entailed by the first explanation, is much more general — it would continue to hold in situations where my efforts to put the peg in the hole were rather different. And it would even continue to hold in situations where the laws were different. If Quantum mechanics were false, and Newtonian mechanics were true, this second explanation would still hold — it would still be the geometric features of the peg and the hole, and their solidity, that explains my failure.

It's natural to ask, then, what is the explanation of my failure to put the peg in the hole, the first or second explanation? Or if both are explanations, which one is better? We will come back to such issues soon. But the point, for now, is simply that there is a big difference between a putative explanation where a phenomenon is explained by appealing to the full details of the situation and an explanation where the phenomenon is explained by appealing to more generic features of the situation. So, an explanation of supervenience that appeals to AU is importantly different from one that appeals to certain generic features of the actual moral principle.

1.1   OPTION 1

Let's consider option 1, where supervenience is explained by AU. The idea is that AU is necessary, and so trivially supervenes on the natural facts. And particular moral facts are necessitated by AU and the natural facts, so they supervene on the natural facts too.

The problem with this putative explanation is that AU is not a *difference-maker* for supervenience. To see this consider, again, Putnam's Peg and the first potential explanation — the one which cites the precise physical details and the fundamental laws. The common judgment, both pre-theoretical and in the relevant literature, is that this is a bad explanation of my failure to put the peg in the hole. A common diagnosis is that this explanation is bad because lots of the things that are cited in the explanans fail to *make a difference* to the holding of the explanandum (e.g. Strevens (2008), Woodward (2010)). The precise physical details of my attempts to put the peg in the hole are clearly not a difference-maker for the fact that I fail to do it.

Exactly what makes one fact a difference-maker for another is a very complicated issue. There are lots of accounts (see Strevens, 2008, section 3.8). But in most cases we have a very good (though not perfect) heuristic that makes it clear when something fails to be a difference-maker — the failure of difference-making is revealed by certain distinctive counterfactuals. In particular, *if I tried to put the peg in the hole in a different way, I still would have failed*. The truth of this counterfactual shows that

the precise way that I tried is not a difference-maker for my failure.[4,5]

Given this, we can see that AU is not a difference-maker for supervenience, because this counterfactual is true:

(**1**) If AU were false then supervenience would still hold.

The truth of this counterfactual is, I think, extremely intuitive. Supervenience doesn't depend upon the truth of AU specifically. If deontology was true rather than utilitarianism then this wouldn't change the facts about supervenience — the moral facts would still supervene on the natural facts.

The issue, here, is complicated by the fact that the standard view is that AU is metaphysically necessary. So (1) is a *counterpossible*. And some argue that counterpossibles like (1) are true, but only trivially, since all counterpossibles are trivial (see Kocurek (2021, section 2.1)).

We will discuss triviality more in section 4. But, for now, note that while (1) strikes us as true this counterfactual strikes us as false:

(**2**) If AU were false then supervenience would not hold.

(2) seems plainly wrong. Nothing about the falsity of AU would mean that supervenience fails. But if counterpossibles are trivial then (2) would be true. Again, we will discuss this more later, but I'll proceed on the assumption that counterfactuals like (1) are non-trivially true.

The non-trivial truth of (1) reveals that AU is not a difference-maker. If we change AU, it doesn't make a difference to the holding of supervenience. Because AU is not a difference-maker, appealing to AU doesn't give us a satisfactory explanation of supervenience.

I think Enoch (2011, p. 149) was getting at something close to this point when he noted that 'the intuitive thought here is that regardless of the true first order theory, there can be no normative difference without a natural one' and, consequently, '*general* supervenience cannot be explained by

---

[4]The reason that this heuristic is not perfect is closely related to problems that are faced by counterfactual accounts of causation – to do, for example, with problematic cases of backups and preemption. The cases that we are focused on in this paper are not of this problematic kind. So we can be confident that such counterfactuals are a good guide to the failure of difference-making in our cases.

[5]This appeal to difference-making is a simplification of the vast literature there is on these issues — particularly connected to the question of how special science explanations of phenomena can have greater, or different, value than competing explanations given by physics. Some authors, instead of appealing to difference-making, appeal to concepts like *proportionality* (Yablo, 1992; Woodward, 2018), *robustness* (Wilson, 1994; Weslake, 2010), or *modal informativeness* (Jackson and Pettit, 1992). The first two are very similar to difference-making, the third less so. But these disagreements don't matter for us, because all of these accounts imply that the first explanation in the Putnam's Peg case is flawed in a way illustrated by the truth of counterfactuals like *if I tried to put the peg in the hole in a different way, I still would have failed*. And this reasoning is what will be important going forward. A rare view in the literature is that the first explanation in the Putnam's Peg case, and similar explanations, are not flawed at all — though perhaps they are harder for us to fully understand. This, I take it, is very unintuitive. It's a commitment of this paper that such a view is wrong.

references to the content of specific norms'.[6] This, kind of idea is best understood, I think, as a thought about difference-making, one that is naturally developed in the way we have done in this section.

Here's one potential concern for the idea that AU doesn't properly explain supervenience – it's that there are some cases where the stipulation of AU seems to make supervenience unmysterious. Consider a case like this[7]: God stipulates what worlds are metaphysically possible, and in particular, they stipulate that AU holds necessarily. This stipulation seems to make supervenience unmysterious – *of course* supervenience holds given this story about God's stipulations. This might suggest that AU does, in fact, explain supervenience.

I agree that this story about the stipulation of AU does make supervenience unmysterious, but that doesn't mean that AU explains supervenience because, again, AU is not a difference-maker. The particular stipulation of AU is not a difference-maker for supervenience. More plausibly, I think, the element of the story that makes supervenience unmysterious is not the particular stipulation of AU but the fact that God stipulated a basic moral principle with the right generic features to guarantee supervenience. As we noted earlier, this is very different from AU – an explanation of supervenience that appeals to AU is importantly different from one that appeals to certain generic features of the actual moral principle. This appeal to generic features of the moral principles is the second option for explaining supervenience, we will discuss it shortly.

1.2 A second challenge

But now that we have seem the argument against option 1, and before we go onto option 2, it will be useful to take a slight detour. The discussion of option 1 might suggest a slightly different challenge that we can raise against the non-naturalist.

I've been discussing the classic challenge for the non-naturalist to explain supervenience. In the last section, I suggested that the non-naturalist can't appeal to AU to explain supervenience because AU is not a difference-maker – supervenience would still hold even if AU did not. Supervenience, it seems, is more *modally robust* than AU. (Again, we are screening off, for now, concerns about the triviality of such counterpossibles.) But putting the point like this suggests an additional striking fact which needs explanation. In addition to the question of what explains supervenience we can ask: what explains why supervenience is so robust?[8]

---

[6]General supervenience being, as we discussed in footnote 2, the type of supervenience that is at issue in this paper, and in most of the literature.

[7]Thanks to a reviewer for suggesting it.

[8]Thanks to a reviewer for suggesting this challenge.

These two challenges for the non-naturalist are interestingly related. The robustness of supervenience consists in (a) the truth of supervenience and (b) certain counterfactuals of the form 'if the moral laws were different then supervenience would still hold'. In the last section, I argued, in effect, that the truth of (b) meant that AU can't explain (a). For that argument, it's irrelevant what explains (b) – that is, it's irrelevant what explains the truth of counterfactuals like 'If AU were false then supervenience would still hold'. As long as such counterfactuals are true, then AU is not a difference-maker for supervenience. But, in addition, if we are trying to explain the robustness of supervenience we face the challenge of explaining why such counterfactuals are true.

My focus, in this paper, is more on the challenge of explaining supervenience rather than explaining the robustness of supervenience. (Though these issues will come up again in section 3.2.) The reason is that, I suspect, the major difficulty with explaining the robustness of supervenience consists in explaining supervenience itself – there may be some relatively easy moves a defender of the divide and conquer strategy can implement to explain counterfactuals of the form 'if the moral laws were different then supervenience would still hold'.

For example, perhaps we can explain the truth of those counterfactuals by saying that the semantics of counterfactuals is such that structural features of laws, and in particular, moral principles, are held fixed across nearby worlds where those principles are false. If that's the case, then that might suffice to explain the relevant counterfactuals, since if AU were false, some moral principles that are similar enough to AU for the divide and conquer strategy to still apply would be true and hence supervenience would still hold.[9] Though, of course, there is a lot more we would have to say to fully evaluate this strategy.

So again, I'm going to mainly focus on the challenge of explaining supervenience and not explaining the counterfactuals that make for the robustness of supervenience, but both are substantial challenges for the non-naturalist.

1.3 OPTION 2

But back to the question of explaining supervenience: If we don't explain supervenience by appealing to the actual moral principle AU what is the other option? Enoch's reaction is to move to a completely different strategy for explaining supervenience — one which focuses on our moral concepts. But we don't need to give up the appeal to moral principles. Rather, we just need to correctly identify the difference-makers. What features of moral principles are relevant for the holding of supervenience?

---

[9] What is it for a moral principle to be *similar enough* to AU for the divide and conquer strategy to still apply? As I'll discuss in the next section it's for those principles to have the form of *bridge-laws*.

As we've noted, the idea of the divide and conquer strategy is that the basic moral principles are necessary, and so trivially supervene on the natural facts. And particular moral facts are fixed by basic moral principles and the natural facts, so they supervene on the natural facts too.

So there are two key points here. The first about basic moral principles being necessary and the second about those moral principles having a particular form, so that all particular moral facts are fixed by the principles and the natural facts. The first point has been the subject of much discussion, following Fine and Rosen's arguments that such principles are not metaphysically necessary — though as we noted we are working with the mainstream assumption that the principles are metaphysically necessary. The second point has had much less discussion. What is needed is for the moral principles to act as something like *bridge-laws* between the natural and moral domains — to take as inputs natural facts and to output the moral facts.[10]

The difference-makers for supervenience, given the divide and conquer strategy, appear to be that (i) the basic moral principles are necessary and (ii) the basic moral principles have the form of bridge-laws between the natural and moral domains.

Three quick clarifications about this: Firstly, there's actually a third point needed for the successful implementation of the divide and conquer strategy – that the particular moral facts are all explained by moral principles. That is to say, the moral principles are *complete* — there are no 'free-floating' particular moral facts that are not governed by any principle. There is an interesting discussion to be had about this notion of completeness, but I'll just assume this point in what's to come. In fact, often when people talk of 'bridge-laws' between domains, they are making analogous assumptions of completeness, so it might be natural to read (ii) as building in this assumption.

Secondly, a note on the right interpretation of (ii): It's not required that *all* the basic moral principles separately have the form of bridge-laws. There could be basic moral principles that connect moral properties. There could, for example, be a basic moral principle that if an action is vicious then it is wrong. But for the divide and conquer strategy to work it must be the case that when we chain the basic moral principles then we get a principle that has a bridge-law form. This is what is needed for the moral facts to be determined by the basic moral principles and the natural facts — as the defender of the bridge-law strategy requires.

Imagine, for example, that in addition to it being a basic moral principle that if an action is vicious then it is wrong it is also a basic moral principle that if an action is vicious if and only if has natural property N. Now the divide and conquer strategy has no problem since we can chain those two moral

---

[10]Don't read too much into this terminology. No assumption is being made that the bridge-laws I'm describing have the same form as those postulated by Ernest Nagel (1979) for example, just that they act as bridges between these domains. Rosen's (2017b) 'Bridge-law non-naturalism' does appeal to moral principles with the form I'm discussing, but he also takes on some further commitments about grounding which aren't needed for our discussion.

principles to generate another moral principle with bridge-law form. So (ii) is correctly interpreted as saying that the basic moral principles, *taken as a whole*, have bridge-law form — so that the set of principles together takes natural facts as inputs and outputs the moral facts. This complexity doesn't make a difference, though, with respect to the example we are focusing on where AU is the only basic moral principle.[11]

Putting together these two clarifications gives us a clearer understanding of exactly what constraint on the basic moral principles a defender of the divide and conquer strategy has to accept. What they need is for the natural facts plus the basic moral principles to entail all the moral facts. This, plus the necessity of the basic moral principles, allows them to say that the moral facts supervene on the natural. That is, they need the moral principles to be such that:

It follows from the moral principles that [for all moral facts F, the moral laws and the natural facts entail F]

From now on in the paper when we talk about the moral principles or laws having bridge-law form we can take that to be shorthand for this constraint holding.[12]

Thirdly, notice that the views we have been considering so far build in (ii). A view where AU is the only basic moral principle satisfies (ii) — it determines what the moral facts are on the basis of natural facts, specifically those about utility.

Further, Fogal and Risberg's picture of how moral principles explain the particular moral facts builds in the assumption that moral principles have bridge-law form.

> **Explanans**: particular natural fact(s) (e.g. a is a lie).
> **Principle**: general explanatory moral principle (e.g. lying is wrong).
> **Explanandum**: particular moral fact (e.g. a is wrong). (p. 175)

Moral principles are, on this approach, explicitly introduced to bridge the natural and moral domains.

A similar thing is true on Scanlon's view. As we noted, for him, moral principles are of the form $\forall x(R(p, x, c, a))$, where this is glossed, roughly, as that if proposition p were true and agent x was

---

[11]Thanks to Daniel Fogal for discussion here.

[12]Some might be worried, though, that this constraint on the moral laws only implies *global supervenience*, not *strong supervenience* which was the type of supervenience we introduced at the start of the paper. If you are worried that global supervenience doesn't entail strong supervenience then you would want a slightly stronger constraint on the moral laws. Here's one way to put it:

It follows from the moral principles that [for all moral properties P possessed by any object x the moral laws and the natural properties of x entail Px]

Whether you think the defender of the divide and conquer strategy needs this stronger condition won't matter for the arguments that I make in the rest of the paper.

in circumstance c, then one reason to do action a would be p. Such principles also have bridge-law form — they take as inputs natural facts — specifically about propositions and circumstances — and output the moral facts about the reasons agents have for actions. It is precisely because of this bridge-law form that Scanlon can use these principles to execute the divide and conquer strategy.

The second option for the divide and conquer strategy, then, is to say that supervenience is explained by the fact that the basic moral principle have features (i) and (ii). This, I think, is a much better option than option one. Given the divide and conquer approach the truth of supervenience seems to depend upon (i) and (ii) in a way that it didn't depend upon AU.

In particular, consider (3), and compare it with (1):

(1) If AU were false then supervenience would still hold.

(3) If the basic moral principles were not metaphysically necessary then supervenience would still hold

As we noted, (1) seems to be non-trivially true. If some other moral theory was true that wouldn't make a difference for supervenience. But the defender of the divide and conquer strategy is not similarly compelled to accept the non-trivial truth of (3). (Remember, we are understanding supervenience as a claim about *metaphysical* necessity.)

The structure of the divide and conquer strategy is such that supervenience depends upon the necessity of the basic moral principles. Again, the idea is that the basic moral principles are necessary, and so trivially supervene on the natural facts. And particular moral facts are necessitated by basic moral principles and the natural facts, so they supervene on the natural facts too. Without necessity of the principles then this strategy clearly cannot work. So, given this strategy (3) looks like it can't be non-trivially true.

(An aside: If, instead of understanding supervenience as a claim about *metaphysical* necessity, we were thinking of it as a claim about *normative* necessity then the metaphysical necessity of the basic moral principles would not be a difference-maker for this sense of supervenience. Rather, the normative necessity of the basic moral principles would be the difference-maker. Either way, the defender of the divide and conquer strategy will think that the necessity of the basic moral principles is a difference-maker for a supervenience principle, formulated using the matching grade of necessity. But in what's to come we will continue to work with supervenience understood as being about metaphysical necessity.)

Similarly, the defender of the divide and conquer strategy is not inclined to accept that this counterfactual is non-trivially true:

(4) If the basic moral principles were not bridge-laws between the natural and the moral then supervenience would still hold

Again, on the divide and conquer strategy, supervenience seems to depend upon the moral principles having this form. It is only because the basic moral principles are bridge-laws that we can say that the particular moral facts are necessitated by the basic moral principles and the natural facts. Without this claim the divide and conquer strategy cannot work. So, given this strategy (4) looks like it can't be non-trivially true.

It's plausible enough, then, that that the divide and conquer strategy can be developed in this second way — since (i) and (ii) seem to be difference-makers for supervenience. There are still places some-one might resist, but I'm happy to concede that supervenience can be explained by the fact that the basic moral principles have features (i) and (ii).

2   Two alternative worries

So, if the divide and conquer strategy can explain supervenience, in line with option 2, then what's the problem? The problem, I'm going to claim, is that even though this strategy can explain super-venience it does so by incurring another, similarly large, explanatory debt. The strategy generates another striking fact that really needs explanation – one that is very closely related to supervenience. Before I describe this concern in detail, though, I want to distinguish it from two other concerns in the area.

One objection to the divide and conquer strategy is that it doesn't avoid accepting necessary connec-tions between distinct existences. Necessary moral principles, like AU, postulate necessary connec-tions between natural properties and sui-generis moral properties, but, the objection goes, we should not accept necessary connections between distinct existences (or at least not happily).

Similarly, Schroeder (2015, p.197) notes that Scanlon's version of the divide and conquer strategy faces this objection (though he doesn't fully endorse the objection). Scanlon's view is that moral principles are of the form $\forall x(R(p, x, c, a))$. However, Schroeder notes that 'to say that R ever holds of any tuple $\langle p, x, c, a \rangle$ is just to say that there are some necessary relationships that hold among wholly distinct entities.' That is, it's to say that there is a necessary connection between the moral reasons agents have for actions and the relevant propositions and circumstances that hold.

But my objection isn't that we shouldn't accept necessary connections between distinct existences. In many cases, it seems reasonable to appeal to necessary connections to explain things that would previously not be explained. For example, as we mentioned earlier, it's common, in the literature on laws of nature, to appeal to necessary connections in order to explain the patterns of events that we observe. Of course, this anti-Humean approach to laws is controversial but it doesn't seem helpful to use the denial of necessary connections as a *premise*.

The supervenience problem for the non-naturalist, as I understand it, isn't specifically about the existence of necessary connections. Rather, the concern is that certain patterns, and, in particular, certain modal patterns like supervenience, are very suggestive of another theory that can properly explain such patterns. The supervenience of the moral on the physical is very striking and in need of explanation for the non-naturalist, not simply because it describes necessary connections between distinct existences, but because of the way that it so suggestive of naturalist theories where supervenience is explained by the moral facts depending upon, or being reduced to, the natural. Consequently, it becomes a challenge for the non-naturalist to explain the phenomenon. Here, for example, is Kim (1998, p. 167) making this point about suggestiveness.

> supervenience itself is not an explanatory relation. It is not a "deep" metaphysical relation; rather, it is a "surface" relation that reports a pattern of property variation, suggesting the presence of an interesting dependency relation that might explain it.

This kind of understanding of the supervenience problem is in the spirit of Leary (2017).

Here's another objection to the divide and conquer strategy that I want to distinguish from my objection. Some might claim that it's bad to leave fundamental moral principles, or pure moral facts, unexplained. For example, Väyrynen (2017, pp. 178-9) in discussing Scanlon's version of the strategy claims 'But for each pure normative truth we can ask: why does R hold of $\langle p, x, c, a \rangle$, not of some other tuple $\langle q, y, d, b \rangle$? The distribution of the R relationship over facts, agents, circumstances, and actions shouldn't be arbitrary.'

There are a few different ways of understanding Väyrynen's point here, but at least part of the concern seems to be that it's theoretically costly to take such pure moral facts as unexplained. He says, 'The issue is what explains why R holds of some tuple when it does'. This is, perhaps, a reasonable concern, — I won't try to evaluate it here — but my concern is not about whether the basic moral principles need explaining.

Of course, there is more to say about both these objections. The key point is just that neither of these objections to the divide and conquer strategy are the one I want to push.

3 ANOTHER EXPLANATORY DEBT

Again, option 1 for the divide and conquer strategy — saying that supervenience is explained by AU — doesn't look promising, since AU is not a difference-maker for supervenience. An explanation that appeals to the specific actual moral principles is unsatisfactory, since supervenience would still

hold even if those principles were different. Option 2 explains supervenience by appealing to the facts that (i) the basic moral principles are necessary and (ii) the basic moral principles have the form of bridge-laws between the natural and moral domains. This looks more promising.

My objection to option 2 isn't, as you might think, about (i). Rather, it's about (ii). The concern is based on two claims: (ii), for the non-naturalist, is a striking fact that really needs explaining. And the divide and conquer strategy doesn't give us tools to satisfactorily explain it.

Let's consider these claims in turn. To start, why think that (ii) needs explaining?

3.1 (II) NEEDS EXPLAINING

Let's consider a few reasons:

Firstly, note that (ii) is extremely modally robust. As we discussed, supervenience is very modally robust — it would continue to hold even if the basic moral principles were different. Specifically, if AU did not hold then supervenience would still hold. (And of course, this isn't specific to AU — in a world where deontology is true then the counterfactual *if deontology did not hold then supervenience would still hold* is also true.)

(ii) is modally robust in much the same way. Someone who accepts explanatory moral principles, will naturally accept (5):

(5) If the basic moral principle was not AU then the basic principles would still have the form of bridge-laws.

That basic moral principles have the form of bridge-laws is not simply because act utilitarianism is true. Rather, every serious candidate for the true set of moral principles has bridge-law form. If the basic moral principle was not AU then some other principles would link the natural and the moral. This robustness of (ii) — the way in which it appears to be extremely non-accidental — invites an explanation of (ii).

Secondly, and perhaps most importantly, (ii) expresses the core idea of supervenience just transferred into a different metaphysical setting. The core intuition of the supervenience of the moral on the natural — the first sentence we say to our students when explaining the concept — is that there can't be a change in the moral features without a change in the natural features. Of course, we commonly formalize this idea using the tools of modal logic — as we did at the start of the paper. And the use of 'can't' in the expression of the intuition very much suggests this type of formulation.

But, if we have the tools of metaphysical laws in hand, then we could formulate the intuition using those tools. The idea would be that the metaphysical laws are inconsistent with there being a change

in the moral without a change in the natural. How would the metaphysical laws have to be for this to be true? Well, the laws would have to say that the moral facts are a function of the natural facts — that the laws would take natural facts as inputs and then uniquely output the moral facts. But that is what (ii) says that the basic moral principles are like.

Supervenience, as we formulated it in this paper, and (ii) get at the same underlying intuition. It's just that supervenience formalizes the intuition using the tools of metaphysical necessity while (ii) formalizes it using the tools of metaphysical law. Of course, it's still possible to claim that, even in light of this, supervenience is striking and really calls out for explanation, while (ii) does not. Perhaps there is a sense in which facts about metaphysical laws, even very robust facts, don't call out for explanation as much as facts about metaphysical necessity. I have some sympathy for this thought. But I don't think it exhausts the puzzle about supervenience. The problem with the the supervenience of the moral on the natural isn't merely that it is an unexplained necessity, so that if you express the same kind of idea using some other metaphysical tool instead of an appeal to metaphysical necessity then the problem disappears. There is a problem with the underlying idea, whether expressed in modal terms or in terms of metaphysical laws — that there is some sort of guarantee that there is no change in the moral without a change in the natural, and this is true even if the actual moral theory was very different. The problem is that there is no reason to have expected this guarantee to hold, given non-naturalist realism — given, that is, that moral facts are sui-generis, causally inert, and wholly distinct from natural properties — but it's to be expected, and is explained, given naturalist theories of morality. This concern still applies to (ii).

Thirdly, just to emphasize a point just made, (ii) is extremely suggestive of alternative theories to non-naturalist realism. That there are principles with a bridge-law form, connecting the natural and the moral is completely to be expected if moral facts are nothing over and above natural facts. (Of course, the metaphysical status of these principles will differ a lot given these different metaphysical views. A moral principle connecting sui generis moral facts and natural facts will be different in nature from principles connecting moral facts which are nothing over and above natural facts.) This is a powerful reason to think that the non-naturalist needs to explain (ii).

Fourthly, the way in which (ii) needs explanation might be hard to see because its hard to see what the other options are. What even would moral principles that are not of this bridge-law form look like? Well, notice that most fundamental properties are not governed by anything like bridge-laws. Take the physical property of charge, for example. It's not governed by anything like bridge-laws between domains. Rather, it's governed by principles that connect it to other physical properties. If moral principles worked like this, connecting moral properties to other moral properties, then (ii) would not hold. If, for example, all the basic moral principles were things like *if an action is vicious then it is wrong*, then the set of moral principles would not have bridge-law form.

To summarize: (ii) is extremely modally robust; it expresses the same intuition as supervenience, just in a different metaphysical setting; it's very suggestive of alternative theories to non-naturalist realism that would explain it; and there are alternative kinds of moral principles that we could write down. These factors are reasons to think that the non-naturalist should explain (ii).

3.2 The divide and conquer strategy doesn't explain (ii)

However, the divide and conquer strategy doesn't give us the tools to satisfactorily explain (ii). The reason is familiar from earlier in the paper. The only candidate that the strategy seems to give us for such an explanation is the actual moral principle, AU. But AU is not a difference-maker for (ii) because (5) is true — that is, if the basic moral principle was not AU then the basic principles would still have the form of bridge-laws. Consequently, AU does not satisfactorily explain why the basic principles have this bridge-law form, just as it does not satisfactorily explain supervenience. The content of the true moral theory does not satisfactorily explain structural facts, like (ii), which would have held even if the moral theory were different.

Similarly, to put it in the terms of Scanlon's theory, the only candidate for explaining why his pure moral principles have bridge-law form is the principles themselves. But again, they are not difference-makers. Even if the principles about what moral reasons we have were different — they would still act as bridges between the natural and moral domains.

Here's an objection to the idea that the divide and conquer strategy can't explain (ii): There's a somewhat natural thought about how (ii) could be explained. AU implies (ii). And there is a natural explanation for (5) — the fact that if the basic moral principle was not AU then the basic principles would still have the form of bridge-laws — it's that worlds where the basic moral principles have bridge-law form are closer to the actual world than ones where the principles do not have this form, since AU has this bridge-law form.[13]

This line of thought is rather intuitive, but there are two problems with it as a defense of the divide and conquer strategy. Firstly, this line of thought still appears to say that the explanation for (ii) is AU. And it's still the case that (5) tells us that AU is not a difference-maker. What the explanation is for (5) is irrelevant to the argument I've been making. Secondly, why is it the case that worlds where the basic moral principles have bridge-law form are closer to the actual world that ones where the principles do not have this form? Of course, similarity matters for closeness of worlds, but why is this aspect of similarity particularly important? Why not any of the other ways in which moral principles could be similar? Saying that worlds where the basic moral principles have bridge-law form are closer

---

[13]Thanks to Daniel Fogal and Olle Risberg for discussion.

to the actual world than ones where the principles do not have this form is just a way of restating the point that I stressed earlier — that (ii) modally robust. It isn't an objection to my approach.

But this suggestion about how (ii) could be explained reveals that again, there are two distinct challenges for the non-naturalist here. In section 1.2 we discussed how there is the challenge of explaining supervenience, which I was focused on, and there is also the challenge of explaining why supervenience is so modally robust – why it continues to hold even in situations where the basic moral principles are different.

Similarly, there is the challenge of explaining why the basic moral principles have bridge-law form, and there is also the challenge of explaining why the bridge-law form of the moral principles is so modally robust – why even if AU didn't hold and some other set of basic moral principles did hold then those principles would have bridge-law form.[14] The modal robustness of basic moral principles having bridge-law form consists in the principles actually having that form, plus the truth of certain counterfactuals like (5) 'If the basic moral principle was not AU then the basic principles would still have the form of bridge-laws'.

Just as my focus was on explaining supervenience, and not on explaining the robustness of supervenience, my focus here is on the challenge of explaining why the basic moral principles have bridge-law form, and not on explaining the robustness of this. The reason for this is similar to the discussion in section 1.2 – as we just noted, it's plausible that we can explain counterfactuals like (5) just by making a claim about the semantics of counterfactuals to the effect that certain structural features of laws, like their bridge-law form, are held fixed across nearby worlds where the laws are false. But regardless of this, there would still be a problem of explaining why the basic moral principles have bridge-law form since AU is not a difference-maker – so this is the challenge I focus on.

(ii), the fact that the basic moral principles have this very specific bridge-law form, is a striking fact that needs explaining for the non-naturalist. But the metaphysical postulates of the divide and conquer strategy don't give us the material we need for a satisfactory explanation. This is a problem with the divide and conquer strategy for explaining supervenience, and more generally, with versions of non-naturalist realism that are built upon a metaphysics of moral principles.

To be clear, the problem is not that supervenience doesn't get explained. (i) and (ii) together explain supervenience. But this explanation of supervenience incurs another, serious and very closely related, explanatory debt: the need to explain (ii). The core intuition of supervenience – the presence of some guarantee that there is no change in the moral without a change in the natural – recurs in (ii) and continues to need explanation.

---

[14]Thanks to a reviewer for discussion of this.

Given that there are alternatives to non-naturalism that can explain supervenience, and the bridge-law form of moral principles, then this is a serious problem for non-naturalist realism.

4 TRIVIALITY OF COUNTERPOSSIBLES

It's time to go back to an objection first mentioned in section 1.1. The objection is that the key counterfactuals I appealed to — things like 'If AU were false, then supervenience would still hold' and 'If the basic moral principle was not AU then the basic principles would still have the form of bridge-laws' — are trivial because the antecedent is metaphysically impossible. And this triviality undermines some of the claims I made using these counterfactuals — like the claim that AU is not a difference-maker for supervenience, or for (ii).

The view that counterfactuals with metaphysically impossible antecedents are vacuous or trivially true flows from traditional accounts of their semantics where a counterfactual tells us about the nearest possible world where the antecedent is true. But this gets trivialized when there is no world, and hence no nearest world, where the antecedent is true.

Notice, as something of an aside, that this reasoning doesn't apply to views that deny that moral principles are metaphysically necessary, instead claiming that they are only *normatively necessary*. Accepting this view makes it very clear that counterfactuals like 'If AU were false, then supervenience would still hold' are not trivial. So this objection certainly doesn't work given such moral contingentist views.[15]

But let's put those views aside again, and continue with views that assume that pure moral principles are metaphysically necessary. The natural way to use the triviality of counterfactuals to oppose my view is to argue that supervenience is explained by AU, the actual moral principle. I claimed that this view was not tenable, because the counterfactuals 'If AU were false, then supervenience would still hold' is true. And so AU is not a difference-maker for supervenience. But, one could claim, this counterfactual is only trivially true, and such trivial truth doesn't imply that AU is not a difference-maker.

There is something initially strange about this view, since, as we noted, this kind of reasoning doesn't work if we are contingentists about morality. It's somewhat strange to think that the moral contingentist has an explanatory problem here — that they can't explain the way that the moral supervenes on the natural in the space of normatively possible worlds by appealing to the the normatively necessary principle AU — but once we claim that AU is metaphysically necessary then we can explain

---

[15]More generally, given our discussion of the supervenience problem, this move of refomulating supervenience as a thesis about normative necessity doesn't really help with the problem (contra Rosen (2017a, section XI)). It's still the case that AU is not a good explanation of supervenience, and it's still the case that (ii) very much needs explaining.

the way that the moral supervenes on the natural in the space of metaphysically possible worlds by appealing to the metaphysically necessary principle AU. This modal difference doesn't seem to have much explanatory import.

But, more substantively, there is a good case to be made that counterfactuals like 'If AU were false, then supervenience would still hold' are non-trivial. Firstly, as we noted in section 1.1 this counterfactual just seems very intuitive, while the counterfactual 'If AU were false, then supervenience would not hold' seems obviously wrong.

Secondly, and closely relatedly, these counterfactuals clearly make sense. We have a clear idea of lots of things that would be the case if AU, or deontology, or some other moral theory held — the literature on those moral theories tells us such things. It's not like asking what would be the case if there were round squares.

Thirdly, there are other important cases of counterpossible non-vacuity. Scientific practice gives us plenty of cases. Consider, for example, 'If diamond had not been covalently bonded, then it would have been a better electrical conductor' (Tan, 2019). This is true, but it's metaphysically impossible that diamond is not covalently bonded. Similarly, scientific modeling very often tells us what would be the case had certain assumptions held, even when the assumptions are metaphysically impossible. For example, the Lotke-Volterra equations in population ecology tells us how predators and prey would interact given certain assumptions — like that populations are continuous — but it's metaphysically impossible that a population of rabbits, say, is continuous. You can't, in the relevant sense, have a population made up of 482.843 rabbits (Jenkins and Nolan, 2012; McLoone, 2021).

Of course, these are not knockdown arguments. The question of counterpossible vacuity can't be resolved here. But there is a lot of pressure against vacuity. The non-naturalist who wants to claim that counterfactuals like 'If AU were false, then supervenience would still hold' are trivial will be pushed to some unattractive commitments.

5   Conclusion

Explaining supervenience by appeal to moral principles seems extremely attractive. But, I've argued, this approach either doesn't give a satisfactory explanation or it incurs a further, substantial, explanatory burden. This approach can explain supervenience but only by committing to a principle that expresses something very close to the core idea of supervenience, just transferred into a different metaphysical setting.

Making sense of supervenience is one of the key challenges for a non-naturalist metaphysics of morality. Views which just appeal to moral principles to explain supervenience look like they fail this

challenge.

But this isn't to say that the non-naturalist has no options here. They do have options for explaining supervenience and, indeed, explaining why supervenience is so modally robust, *if* they bring in other metaphysical machinery. For example, a reviewer suggests a possible strategy which is centered around appeals to *essence*. The idea is that the non-naturalist claims that it's in the essence of moral properties like rightness and wrongness that they are explained by natural properties, and it's in the essence of the relevant type of explanation that what it is for the natural to explain the moral is for there to be some general moral principles linking the natural and the moral.

This strategy would give us an explanation of supervenience from the facts about the essence of moral properties. And if we assume that the facts about essence are held fixed even in worlds where the moral principles are different then we also have an explanation for why supervenience is so modally robust – why it continues to be true even in cases where AU is false. But what's more, given the claim that it's in the essence of the relevant type of explanation that there are moral principles linking the natural and the moral, this is a picture where moral facts are governed by moral principles.[16]

This is a very elegant strategy. It would take a lot more discussion to fully evaluate it, but it has some initial plausibility. Notice, however, that the work in explaining supervenience, and in explaining the robustness of supervenience, is done by the appeal to the essences of moral properties. Similarly, the non-naturalist has other possible strategies if they allow the work to be done by an appeal to essence. However, as I've been arguing, versions of non-naturalism that attempt to explain supervenience simply by appealing to moral principles fail to meet the challenge. The non-naturalist should look to other metaphysical machinery, like that of essence, to develop their view.

References

Armstrong, D. (1983, January). *What is a Law of Nature?* CUP.

Atiq, E. H. (2020, September). Supervenience, repeatability, & expressivism. *Nous 54*(3), 578–599.

---

[16]As the reviewer notes, this strategy has similarities in spirit to Wedgwood's (2007) discussion of supervenience. Wedgwood's view also suggests that supervenience would still hold even if the true moral theory were different. This comes out in the way that Wedgwood takes the true moral theory to be only contingently necessary, while supervenience is necessarily necessary. Developing this view involves being somewhat revisionary about the logic of metaphysical necessity. He rejects S5 – specifically, rejecting the axiom that if a truth is necessary then it is necessarily necessary.

This strategy is, I think, a little less plausible than the version of the strategy discussed in the main text because of the somewhat revisionary modal framework. And I think there isn't a lot of motivation for defenders of the divide and conquer strategy to take the Wedgwood approach because it's motivated by the idea that our explanations of things like supervenience have to bottom out in facts about essence plus contingent facts (Wedgwood, 2007, pp. 207-208, 210). But it's much more natural, once we have a substantive notion of moral law or moral principle in hand, as defenders of the divide and conquer strategy do, to allow that such explanations can bottom out in facts about essence and about moral principles.

Bhogal, H. (2020, August). Humeanism about laws of nature. *Philosophy Compass 15*(8), 38.

Elliott, A. (2014). Can moral principles explain supervenience? *Res Philosophica 91*(4), 629–659.

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

Fine, K. (2002). The varieties of necessity. *Conceivability and possibility*, 253–281.

Fogal, D. and O. Risberg (2020). The metaphysics of moral explanations. *Oxford Studies in Metaethics Volume 15 15*, 170.

Jackson, F. and P. Pettit (1992, January). In defense of explanatory ecumenicalism. *Economics and philosophy 8*(1), 1–21.

Jenkins, C. S. and D. Nolan (2012). Disposition impossible. *Noûs 46*(4), 732–753.

Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-body Problem and Mental Causation*. MIT Press.

Kocurek, A. W. (2021, November). Counterpossibles. *Philosophy compass 16*(11).

Leary, S. (2017, July). Non-naturalism and normative necessities. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaphysics Volume 12*. Oxford University Press.

McLoone, B. (2021, December). Calculus and counterpossibles in science. *Synthese 198*(12), 12153–12174.

Nagel, E. (1979). *The Structure of Science: Problems in the Logic of Scientific Explanation*. Hackett.

Putnam, H. (1975, January). Philosophy and our mental life. In *Mind, Language and Reality: Philosophical Papers, volume 2*. Cambridge: Cambridge University Press.

Robinson, L. (2011). Moral principles as moral dispositions. *Philosophical studies 156*(2), 289–309.

Rosen, G. (2017a). Scanlon's modal metaphysics. *Canadian journal of philosophy 47*(6), 856–876.

Rosen, G. (2017b). What is a moral law? *Oxford studies in metaethics 12*, 135–159.

Rosen, G. (2020). What is normative necessity? In *Metaphysics, Meaning, and Modality*, pp. 205–233. Oxford University Press.

Rosen, G. (2021). The modal status of moral principles. *Oxford Studies in Metaethics 16*, 257–279.

Scanlon, T. M. (2014). *Being Realistic about Reasons*. OUP Oxford.

Schroeder, M. (2015, January). Being realistic about reasons, by T.M. scanlon. *Australasian journal of philosophy 93*(1), 195–198.

Strawson, G. (1989, January). *The Secret Connexion: Causation, Realism, and David Hume*. Oxford University Press.

Strevens, M. (2008, January). *Depth: An Account of Scientific Explanation*. Harvard University Press.

Tan, P. (2019, March). Counterpossible non-vacuity in scientific practice. *The journal of philosophy 116*(1), 32–60.

Väyrynen, P. (2017). The supervenience challengeto Non-Naturalism.

Wedgwood, R. (2007, July). *The Nature of Normativity*. Clarendon Press.

Weslake, B. (2010). Explanatory depth*. *Philosophy of science 77*(2), 273–294.

Wilson, R. A. (1994). Causal depth, theoretical appropriateness, and individualism in psychology. *Philosophy of science 61*(1), 55–75.

Woodward, J. (2010, January). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy 25*(3), 287–318.

Woodward, J. (2018, October). Explanatory autonomy: the role of proportionality, stability, and conditional irrelevance. *Synthese*, 1–29.

Yablo, S. (1992). Mental causation. *The Philosophical review 101*(2), 245–280.