

Explaining Supervenience

Mark van Roojen

Abstract

It is widely agreed that nonnaturalists incur a burden to explain how it is that the normative or moral must supervene on the nonnormative or nonmoral. This paper argues that every metaethical theory, including naturalist theories, must carry that burden in one way or another. Along the way it notes some troubles for making the naturalism/nonnaturalism distinction precise and relatedly for formulating the supervenience thesis to be explained. It then explores naturalistic options for explaining supervenience, suggests that some of these are better than others, and argues that parallel nonnaturalist theories can generate parallel explanations to those offered by naturalists.

1. Introduction

A popular and plausible take on the advantages of naturalism over non-naturalism is that nonnaturalists incur a burden to explain how it is that the normative or moral must supervene on the nonnormative or nonmoral. (Dreier 1992, McPherson 2012, 2015, Elliot 2014, Bader 2017, Faraci 2017, Leary 2017) There is some disagreement over the nature of the must here – about whether it is the must of metaphysical necessity or a somewhat weaker must of moral or normative necessity (Fine 2002, Rosen 2010 and 2022, Dreier 1992 and 2019), and also over the status of the overall claim - is it a conceptual truth or something weaker? But these disagreements aren't important to my main points. Theorists of all stripes will want to be able to tell a story about how moral and normative properties fit into the environment that presents itself to us in space and time, and to the supervenience relations, whatever their status, that hold between these properties and the natural properties that present themselves to us in that environment. The claim is that naturalists have no similar problem. The standard view seems to be that nonnaturalists bear a burden in this regard that naturalists do not. If this is true, this is a serious issue for the nonnaturalist. There has been a lot of literature focused on the viability of various nonnaturalist responses. (Elliot 2014, Bader 2017, Faraci 2017, Leary 2017, Väyrynen 2017, Toppinen 2018, Fogal and Risberg 2020, Morton 2020,

Bhogal 2022.) But there has been less focus on the ease with which naturalists can do the relevant explanatory work. In this paper I want to focus on that, and I will argue that at least some versions of naturalism will wind up with explanatory burdens not much different from those that face non-naturalists.

To oversimplify my thesis a little bit so as to get it into a slogan: Just because one property is in the same metaphysical category as another does not yet explain why an instance of the first can necessitate an instance of the other property. But I have several more specific theses as well. Naturalism comes in several varieties and different varieties will have different strategies available to explain the relevant supervenience claim. So different varieties will face different challenges, and some will likely have more success than others. The pages that follow will try to take this variation into account and explain which views are most promising. Finally, I will point out that some of the successful strategies could be mimicked by nonnaturalists.

2. We should interpret the disagreement between naturalists and nonnaturalists as fundamentally metaphysical

If the definitive disagreement between naturalists and nonnaturalists is to have any content, there must be some commitment that unites naturalistic metaethical positions and divides them from nonnaturalist positions. This can be hard to find, insofar as naturalism comes in different varieties. Reductive naturalism suggests normative and moral properties can be reduced to other properties, but divides over whether reduction requires a symmetrical relation such as identity (Dorsey 2016, Sinhababu 2018) or whether it requires only an asymmetric relation such as constitution, composition, grounding or another making relation (Schroeder 2007). Nonreductive naturalism looks a lot like old fashioned non-naturalism and seems to differ with it only over whether normative properties are themselves natural or non-natural. Both nonreductive naturalists and non-naturalists think that normative properties are a distinct class of properties from the non-normative. They also both think that the distribution of normative properties supervenes on the distribution of non-normative properties and, furthermore, that the distribution of the non-normative properties in some sense determines the distribution of normative properties. Reductive naturalists in fact agree with these last commitments regarding supervenience. The two sides divide just over whether normative and moral properties are natural or not.

But what does that come to? It would be easy to say if one could give a set of necessary and sufficient conditions for a given property's being natural.

Explaining Supervenience

But no one has provided one with any real bite. This doesn't mean there can't be a substantial disagreement but it does make it hard to say exactly what it is about.¹ Each side of this debate thinks there is some good point to characterizing certain properties and relations as natural and certain others as not. In addition, the parties agree that there is a substantial and interesting question about whether normative properties satisfy the criteria for naturalness that make the distinction worth making. For the question to be substantial and interesting, there must be some unity to the property of being natural or to the property of being non-natural. To put it another way, being natural and being non-natural must be higher order properties of properties, and these higher order properties must constitute a real similarity shared by the properties in their extensions. For a property to be natural or to be nonnatural should not just come to being a property on some list of properties with nothing in common besides being on the list. (Suikkanen 2010, van Roojen 2015, forthcoming, McPherson 2015, 2018) Furthermore, the properties in the extension of these higher order kinds must themselves involve some sort of real similarity so that there is some genuine universal or elite kind² to actually have the higher order property of being natural in virtue of its own nature. Otherwise the higher order property of being natural would not pick out a similarity shared by the natural properties.³ The disagreement, so conceived, comes down to the question of whether there is a higher order similarity among the natural properties in virtue of which they are natural and, if there is, whether this similarity is shared by the normative properties. Naturalists answer yes and yes. Nonnaturalists answer yes but no.

While it would be nice to have a more substantial characterization of the natural or non-natural, we can get some mileage out of this more abstract way of characterizing the dispute.

¹I've expressed some doubts about whether we can in fact make a division that both captures a substantial distinction between the positions and tracks the arguments made in support of each side in van Roojen (forthcoming).

²David Lewis (1983) following David Armstrong (1978) distinguished between abundant properties and genuine universals and sometimes used the term 'natural' to refer to the universals. Since this was a kind of pun it would be confusing to use that terminology here. Billy Dunaway and Tristram McPherson (2016) use the term 'elite' for the genuine universals, and various theorists have pointed at the same or a similar divide by using "real similarity", "joint cutting", "non-Cambridge" as modifiers to pick out the elite side of the division. I sometimes use each of these terms in this paper and I mean to use them synonymously though it could turn out that this is a mistake.

³A slight complication: It could be that the non-natural properties form an elite kind of elite kinds, and that the property shared by the natural properties is not-being-one-of-those. I guess that would be OK. But one or the other of the natural or non-natural should be elite.

3. The Supervenience Claim to be Explained

The supervenience claim that metaethicists generally take to need explanation is that the extension of normative/moral properties supervenes on the distribution of the (seemingly) nonnormative or nonmoral properties. To say that one property supervenes on some others is itself to say something about what can and cannot happen – roughly, there can be no difference in the having or not of a supervening property within some domain without those things also differing in their having or lacking the subvening property. Or, comparing the distribution of properties in one possible world to another, there can be no difference in the distribution of the supervening property between two worlds without a difference between them in the distribution of the subvenient or base properties. The idea can be extended to sets of properties of a certain type. The distribution of properties in one set will supervene on the distribution of the properties in the other set if there can be no change or difference in the distribution of the former set without a change or difference in the distribution of the other set. So the overall claim is that necessarily any change/difference in the distribution of normative properties necessarily requires a change/difference in the (seemingly) other properties.⁴ Some think this is true as a matter of conceptual necessity – assured by the nature of normative or moral concepts (Dreier 1992) – whereas others go for something slightly weaker such as that it is an a priori truth, conceptual or not (Kramer 2009, Scanlon 2014). And most theorists think the necessity of the connection between the two sets of properties is metaphysical necessity, though some non-naturalists (Fine 2002, Rosen 2010) have posited a somewhat weaker sort of necessity that they contrast with metaphysical necessity and call moral or normative necessity. The parenthetical “seemingly” in the above formulations is there to keep from ruling out identity between the supervening properties and their base properties. For it could turn out that the reason the distribution of base properties must change in order for the distribution of the supervening properties to change is just that they are just the same sets of properties. This last point is important insofar as it allows property reductions conceived of as requiring identity to explain supervenience relations. A property might supervene on the arrangement of particles in an object because that property just is a certain

⁴The usual way to state this is just to say that the normative properties supervene on the natural properties. I’m being a bit cagey about the exact set of properties in the supervenience base due to complications needed to leave open the possibility of the normative properties turning out to be identical with some members of the set that makes up the base while at the same time not ruling out supernaturalism as a metaphysical possibility. The two pages that follow explain the problem in more detail.

arrangement of the particles in that object. And so on.

It is actually hard to state very precisely which supervenience claim is both (1) a priori or conceptually necessary and (2) stands in need of explanation if one does not want to rule some positions out by one's formulation. I'm tempted to say that the distribution of the normative properties (including the moral properties) supervenes on the distribution of the rest. Or to put it another way, the world could not be non-normatively as it actually is while having a different configuration of normative properties. If the murder of some person is wrong in the world as it actually is, a murder with all of the same non-normative properties of the first would also be wrong. But this formulation won't quite do if we want to leave open that wrongness might be one of the properties we meant to include in the non-normative base properties (as we should if we don't want to exclude reductive naturalism conceptualized in terms of identity) as a possible explanation. (Sturgeon 2009) The normative property wrongness won't be an "other" property. But it also won't do to pick out some class of properties as natural and then to say that the moral properties supervene on the natural properties. Those who think there can be supernatural properties should probably also think that the distribution of these properties could change the distribution of normative properties. And here it is the metaphysical possibility of such properties that matters to the supervenience claim, not their actual instantiation. (van Roojen 1996) So I am going to work with this: the distribution of normative properties (including the moral properties) supervenes on the distribution of what, for all we know, could well be the other properties, though things might not be as they seem.⁵

4. How might naturalists explain this kind of supervenience?

4.1 Reduction as Identity

I've just emphasized that things might not be as they seem – if naturalists are correct we might not really have distinct sets of properties. The normative properties might just be among the natural properties so it would be no surprise if the normative properties supervened on the natural properties. For

⁵I grant that "for all that we know" is a bit of a weasel phrase. A helpful referee points out that my thought can be read as introducing a kind of relativity into the supervenience base if people differ with respect to their epistemic situations. And certainly, committing to a particular metaethical reduction might put someone in a different epistemic position with respect to a property than someone like myself who lacks any such commitments. I don't intend any such relativity. I am intending to exploit what I take to be our present collective epistemic situation. As far as even the community of philosophers is concerned there is no property we collectively know to be identical to rightness, or impermissibility. And that remains true even if some of us accept some specific identity claim in the way philosophers sometimes do.

they would be among the natural properties and any set of properties supervenes on itself. Jamie Dreier (1992, 2019) has emphasized this and he is correct.

But we need to be careful about how far this gets us. To say that the supervenience claim is a priori is to say we are in a position to know it without any experiential evidence. So, somehow or other, I am justified in believing that the normative properties supervene on the seemingly nonmoral properties. This could be true because unbeknownst to me each normative property is identical to one of the properties I did not take to be normative or it could be true because these normative properties are distinct from each but covary with these possibly other nonnormative properties. For all that most of us are in a position to know a priori, both options are open. We might find out that the identity claim is true. But if we do, the evidence is likely to include empirical elements. So even though identity entails supervenience the identity of a moral property with some base property would not, without more, explain why supervenience is a priori knowable.⁶

When people say that it is conceptually necessary or a priori that the normative supervenes on the non-normative they often defend the claim by pointing out how absurd it would seem if someone pointed to two seemingly identical actions or policies and calling one right while denying that the other was right without being able to point to some other difference to explain the normative difference. And generally we expect them to cite some difference that could conceivably matter morally, though there will be substantive disagreements over which differences can matter. Even when we disagree substantively we will generally recognize that certain differences are sufficient to elude the charge of conceptual incoherence or of making the a priori mistake that would be involved in denying supervenience. I may not myself be a theist and I may think divine commands would carry no real authority, but if my interlocutor cites a supernatural difference as grounding the normative difference I shouldn't accuse them of conceptual confusion or of failing to recognize an a priori truth about supervenience.

My suspicion here is that we know something a priori that entails supervenience but goes beyond the bare supervenience claim. We are somehow aware of a relation of dependence short of identity that explains covariation of certain sorts between normative and nonnormative matters.

⁶There is a bit of trickiness here about what we should say about knowing of each moral property that it is identical with itself. This is a general puzzle about how to speak of our knowledge of identity claims. We all knew that water is water, so if water is H₂O and Leibniz's law holds, in some sense we already knew that. But this doesn't impugn the thought that it was a discovery that water is H₂O, and that we did not always know it.

Explaining Supervenience

Competent users of normative notions have some sense of the kinds of differences that can make for higher-level evaluative differences and we are used to relying on such differences in deciding what is right and wrong. As Kieran Setiya puts it, “The core idea of supervenience is that ethical differences turn on differences of other kinds.” (Setiya 2012, p. 4) We take ourselves to know that certain sorts of relatively granular facts can make a difference to the moral and normative status of individual actions. We are aware that we run into new practically relevant differences all the time and so we don’t take ourselves to be in a position to list all of them. Furthermore, in relying on the properties tracked in this way we are taking their presence or absence to justify certain normative verdicts. As Simon Blackburn suggests, normative judgements are for reasoning practically based on information about our situation and allowing our overall normative verdicts to vary without any difference in situation would undercut the point. (1984, p.186) The identity of a given normative property with some property not obviously connected to facts about the distribution of these somewhat granular seemingly relevant properties won’t really explain all of what we were hoping to capture – at least not without supplementation to connect the fact identified with these putatively relevant facts.

Here is another way to try to get at this. If you persuade me that moral rightness was some complex natural property (and you could consistent with the story so far), I would still think it worth asking why this property covaried with the instantiation or noninstantiation of the supernatural properties which are not instanced in the actual world. This burden could, I think, be discharged. But it isn’t discharged just by convincing me of the identity. I need to know why that property – now identified with some natural property – would supervene on the set including supernatural properties but not on a set which omitted them. Relatedly, if the distribution of moral properties could not change the supervenience claim would be true. Someone might thus offer the following explanation: supervenience holds because all truths about the distribution of moral properties are necessary. But we would find it obviously inadequate. We know the distribution of moral properties is not necessary – part of the point of making moral judgements is to think about the effects of our actions on the distribution. Against this background, saying that the distribution of moral properties supervenes on the distribution of some other set of properties is meant to communicate that a certain dependence relation holds between the possible distributions of properties in one set and those of the properties in the other, and that in order to change the distribution of normative properties you must effect changes in the distribution of the rest. That is the claim that we need to explain.

If that is our situation, everyone will need to explain how that dependence works. It might turn out that some seeming dependence relations are actually identities, so that some of what had seemed to be covariation is really just variation in a single property that had appeared to us to be two different properties. But that will leave us in need of an explanation of why the distribution of some of the remaining properties matters to the distribution of the property so-identified. Suppose it turns out that rightness is just goodness promotion and goodness is constituted by happiness. That would entail supervenience, but we would also want an explanation of why certain differences between actions effected their deontic status. Of course we would have resources to work with given that we do think happiness can be promoted and thwarted in many ways depending on the causal relations between it and various actions. There would be no great mystery here. But that is because we already know a lot of what we would need to in order to construct the full explanation.

One candidate identity which would all by itself quiet the demand for an explanation of supervenience is one which identifies the target normative property with a disjunctive property constructed out of each of the different ways a thing can be consistent with instantiating that normative property. Two actions cannot differ in their rightness without also differing in the presence of absence of at least one of the properties disjoined just because the rightness is a disjunction of just those disjuncts. Unfortunately for the naturalist, this identity won't vindicate naturalism of the sort motivated by anti-supernaturalism since some of the disjuncts here will need to cover the nonactual possibilities where supernatural properties are instantiated and matter normatively. (van Roojen 1996, 2006)

Even apart from that, given the highly disjunctive nature of the base property, there is some pressure not to identify it with the moral property that supervenes on it. At least there is on the assumption that the moral properties are themselves genuine universals tracking real similarities. A disjunction does not convey what the disjuncts have in common. Take any set of specific properties you like and disjoin them. If we ask what these properties have in common such that having any one of them suffices for having a property constituting a genuine universal, it doesn't give any explanation at all to say they all share in the property of being this, that, or the other thing. For we could make parallel claims of any random collection of disjoined properties whether or not the individual instances were sufficient for ascribing a real similarity to anything that instantiated them. Only when the disjoined properties have something in common do things with those properties share a real similarity and that real similarity will be what the term for the property

we are trying to reduce ascribes to the things of which it is predicated. The upshot, I think, is that theories which vindicate naturalism by identifying target normative properties with natural properties will need to identify them with relatively unified natural properties that display these normative properties as sharing some real similarity with the other natural properties.

None of this is to say that naturalism cannot explain supervenience, nor even to say that versions which aim for reductive accounts that identify normative properties with natural properties are not up to the task. The identity of the normative property with a natural property can be at a higher level than the supervenience base. (Sinhbabu 2018) So long as an explanatory story connecting the natural property specified as identical with the target normative property can be connected in an appropriate way with the components of the possibly complicated disjunction characterizing the supervenience base, naturalism can meet the explanatory burden. This will often be possible where the nature of the higher level property requires some sort of role that instances of that property must be able to play and where the properties in the minimum supervenience base can in fact all play that role. (I'll sketch an example of this sort later on in the paper.) To give such an account is to consider the reduction itself to be an identity between one normative property and one unified natural property. But this single property will then stand in some sort of asymmetric dependence relation with the properties in its minimal supervenience base. Or at least it will if the supervenience relation to be explained is a genuine kind of covariation, as I suggest above. The next model considers such asymmetric dependence itself to be what reduction requires while denying any identity between the reduced property and what it is reduced to.

4.2 Reduction as an Asymmetric Dependence Relation

Theorists who think of reduction as constituted by an asymmetric dependence relation rather than as an identity can offer disjunctive naturalistic reduction bases without flouting the demand for real similarities among the things they disjoin. Since the reduced property is not identical to the disjunction that is sufficient for having it, it can constitute a real similarity even while the property of having one or another of the more specific properties in the disjunction does not count as such a similarity. For they are not strictly speaking one and the same thing and hence Leibniz's law doesn't require us to think that everything true of the one must be true of the other. The reduced property will represent the real similarity shared by the property in question, while the disjunctive reduction base can catalogue the diverse ways that things can come to share in that real similarity. For the same reason, having a

reduction base composed of only natural items won't automatically vindicate naturalism. For even if we think that a property is natural just in case it is either a natural elite property or complex property constructed out of elite natural properties via conjunction and disjunction, we can't say that the naturalness of the property offered as the reduction must also be attributed to the reduced property. For this kind of reduction doesn't identify the reduced property with its reduction base. That's the point of saying the relation is asymmetric. So the inference that the property so-composed must itself be natural is a further step. For our purposes we can ignore this issue. We are exploring what naturalists can say to explain supervenience of the normative on the non-normative. To investigate that issue we can just stipulate that the reduced property counts as natural and then ask how what sort of explanation of supervenience such theories can give.

Some theorists suggest that asymmetric making relations such as composition and constitution admit of no explanation.⁷ You might think this would help the naturalist working in this vein. They need not provide an explanation because no such explanation is possible where composition or whatever are concerned. But if it would, it seems like the same move would be available to non-naturalists. The naturalist and the non-naturalist could each offer a constitution story about the target normative property and say their work is done since no explanation of why these concatenations of lower-level properties make for the instantiation of the target normative property.

⁷It is not always clear whether a given theorist is committed to a no explanation view, partly because many authors distinguish metaphysical explanations from "epistemic" or other sorts of explanations and consider metaphysical explanations to just be grounding relations. (See Bennett 2017, pp. 61–2) At the level of metaphysical explanations Bennett (2017, pp. 190 ff.) seems to distinguish 3 views: Suppose A grounds B. Primitivists hold that nothing grounds this fact. Anti-primitivists hold that something does ground this fact and they divide further into two kinds. Upwards anti-primitivists think that A grounds the fact that A grounds B. The remaining kind of anti-primitivism must hold that something further grounds the fact that A grounds B. But this seems to lead to a regress, so Bennett argues for upwards anti-primitivism.

But now think of how this is going to work when role properties are instantiated in virtue of some ground. I think it is explanatory to say that the role property is instantiated because some more particular property instance is capable of playing that role. There is a heart here because there is a muscle connected in the right way to pump blood within an organism. This is obviously incompatible with primitivism. Is it compatible with upwards anti-primitivism? It looks like it is, since we are citing further facts about roles to explain why the particular realizer of that role grounds an instance of the role property. But really it depends on whether we consider those facts to be part of the ground or whether we leave them out. If they are included in A along with the fact that some realizer property is instantiated, then they can fully ground B – that the supervenient role property is instantiated. That way of implementing upwards anti-primitivism would be compatible with offering explanations.

Explaining Supervenience

In any case, I'm not persuaded that asymmetric reductions must always be brute. One reason I am not is that I can think of asymmetric making relations where we can in fact say that there is a further explanation that can be given. In particular, I think that multiply realizable functional properties are grounded in various constellations of more fundamental properties but the nature of the grounded property together with the causal powers of the constellations that ground them together explain why the lower-level properties ground the functional properties. I think that mental properties likely stand in this relation to physical properties and that certain social properties – being money say – stand in such relations to physical and psychological properties.

In fact, one good reason to go in for asymmetric reductions rather than identities depends on relations that work like this. We think that it is essential to certain sorts of mental states such as beliefs or desires that they be able to play certain sorts of roles. We also think that the physical states of certain creatures are capable of playing those roles, perhaps because of their causal powers, or whatever. But we also think that there could be other physical states that also could stand in the specified relations to other counterpart states. These would be playing what is basically the same role relative to their counterparts as the actual physical state plays relative to its partner. Given that there is more than one way a state could play the specified role, we should identify the state with neither of the lower level states that can play the role. What makes the two lower level states still instances of the more general mental kind is the role that they play. Their aptness for playing this role is the real similarity shared by states of that kind.

The example used mental states and their relations to physical states. But similar relations can hold between properties. Obviously so, since we can construct properties such as thinking that it is raining out of states such as states of mind. If a certain constellation of properties constitutes the wrongness of one action and a different constellation constitutes the wrongness of another, the wrongness is multiply realizable by properties at that level. In effect this is just the situation that made problems for reduction as identity view when the reduction base is disjunctive.

This doesn't yet show that there can't be a single unifying specification of wrongness in natural terms. For there might be a real but reducible similarity at the higher level even while the extension of the property has a disjunctive base with each disjunct sufficient for necessitating (and hence grounding) the higher level property. Views like this could then be accepted consistent with reduction as identity. But on the other hand, there might still be other reasons to stick with reduction as an asymmetric relation. In any case, a higher level

similarity shared by the lower level disjuncts in the supervenience base would put such naturalists in a position to explain supervenience. I'll sketch two strategies for constructing a naturalistic theory where the real similarity shows up at a higher level, and suggest that these would be well-placed to explain supervenience. That such strategies are available shows that naturalism can meet the challenge I am posing and also that the challenge is not unfair. From there I will explore whether non-naturalists might be able to give parallel explanations, and if so what this all shows.

5. Strategies of Naturalistic Reduction that can Explain Supervenience

5.1 First Model: Moral Functionalism

One model for this in ethics is moral functionalism of the sort proposed by Frank Jackson and Philip Pettit. (1995) We specify a role that each moral property has to play, possibly interdefining various moral properties, and then look for a property which plays each of the roles. Jackson and Pettit develop this as an identity theory but you could use a similar rationale to generate an explanation of asymmetric dependence on a set of base properties by a higher and multiply realizable normative property. To illustrate with an oversimplified toy example, suppose you thought that moral property M must be able to motivate rational agents under certain conditions and that this was constitutive of property M. And you thought various configurations of other properties N1, N2, N3 . . . , were such as to be able to do that so that wherever one of those was instantiated appropriate motivation followed.⁸ A disjunction of those properties might then be an appropriate reduction base for the property M. We would have a two part explanation for why property N1, N2, N3, etc. form the reduction base for M. One part would tell us that it was constitutive of M that it be able to play a certain complex role which includes being able to motivate actions in certain conditions. The other part would note that instances of each of the properties disjoined in the reduction base are able to play that motivating role.

This last part would give us a perfectly good explanation of why the distribution of M supervenes on a base consisting of N1, N2, N3, and so on. The explanation is parallel to an explanation of why facts about what is a heart and what is not supervene on the physical arrangement of component parts and perhaps also on connections with bodies that require the pumping of fluid through these bodies. Hearts are defined by their role of pumping fluids through the body of living things, perhaps in certain sorts of

⁸My oversimplification ignores most of the network that defines the property in question, including its relation to other target normative properties that are also explicated via the analysis. These are important to the Jackson and Pettit proposal but not relevant to the issue at hand.

environments. Ordinary organic hearts and artificial hearts are both enabled to do this kind of thing because of their physical makeup. A device or organ that could not be and could never have been employed to do this kind of thing due to its construction will not be a heart. So it is reasonable to think that hearts have to be realized in some sort of physical stuff, that being a heart must supervene on properties of this stuff and its arrangement. The explanation involves both the nature of the supervenient property that involves being able to do a certain thing or things and the nature of the various subvenient properties that are capable of composing something that can play those roles.

So long as the description of the role was genuinely naturalistic and unified, there is a good sense in which this would be a version of naturalism. Every instantiation of the property would share in this real higher level natural similarity. And this would be true even when a thing's having that property in a particular situation might depend on supernatural facts at a lower level of description. Of course it may be difficult to say whether the functional property specified by the network for any given normative property is in fact natural, both because the network is complicated and because we at present lack a precise substantive characterization of the natural. But on the assumption that such a characterization is in principle available, this sort of view could vindicate naturalism. And it could explain supervenience by showing how each of the lower level realizers in the supervenience base in fact is sufficient for necessitating the higher level natural property by playing the role specified in the Jackson and Pettit analysis.

5.2 Second Model: Actualist Relationalism

Another naturalistic model able to explain supervenience even on supernatural stuff is relational and actualist. According to such views moral properties are constituted by relations to actual natural entities, most plausibly agents or observers. The basic schema I have in mind is to reduce normative properties to relational properties, where one relatum is an agent or observer or type of agent or observer. Ideal observer views of certain sorts, ideal agent views, and variants of views like that of David Lewis's (1989) analysis of value will all qualify. This kind of theory makes it easier to decide whether the property we have identified with our target property is natural than does a network analysis like that of Jackson and Pettit. That is because it is simpler. And with simpler properties it can be easier to be confident that they are natural without having a definition of naturalness in hand. If the nature of the observers or agents can be specified in fully

naturalistic terms and the relevant relation to them is also natural, then the relational property that is constituted by being so-related to an agent of that kind will also be naturalistic.

An example is Roderick Firth's (1951) version of the ideal observer theory. Firth tells us that actions are right iff they would meet with the appropriate sort of approval (a naturalistically characterizable psychological state) from ideal observers of a certain sort. He lists a bunch of characteristics of such ideal observers, including being fully informed both about the objective facts (omniscience) and about the character of people's experience of those facts (omnipercipience), lack of bias, and consistency. Each of these is arguably a trait that can be thought of as a natural or naturalism-compatible property of persons.⁹ And then he goes on to say that the relevant ideal observer will be "otherwise normal". This last clause is crucial, for how an observer of a sort characterized by the other properties already listed would react is underdetermined without more constraints. Psychopaths might not react positively to paradigmatically right actions. We have no idea how omniscient space aliens might react. Or for that matter omniscient and omnipercipient unbiased consistent ferrets. By telling us the relevant observers are otherwise normal, Firth is ostensibly defining the rest of the observers' psychology. Ideal observers have the psychologies of actual normal humans, whatever those turn out to be. Firth doesn't completely spell this out but, if we are trying to figure out the status of an action undertaken in a world with only psychopaths, we don't look to the reactions of omniscient, omnipercipient, unbiased consistent psychopaths. Rather we ask how people with all of those characteristics but with normal psychologies would react.

This allows us to say that actions by supernatural creatures could be right and could be wrong without building anything supernatural into the nature of rightness or wrongness itself. To call an action right is just to say that normal

⁹Readers may worry that ideal observer theories are not naturalistic, possibly because of the use of the term 'Ideal' in the name. I think some are and some aren't. I think it depends on the particular characteristics built into the analysis, and also on what makes a property natural. Firth's observer's characteristics seem to me to be good candidates for naturalness, at least so long as psychological properties count as natural. The naturalist will likely agree, since one motivation for naturalism is the thought that all instantiated properties are natural, and ordinary people have less extreme versions of the properties listed (some knowledge, rather than omniscience, some empathetic understanding of others rather than omnipercipience, etc.) though not to the degree that the ideal observer is supposed to have them.

Nor should the name mislead us. 'Ideal' in 'Ideal Observer' functions as it does in 'ideal plane' when we encountered it in grade school physics. It references an idealization of a property or set of properties taken to their limit. Thanks to a helpful reader who urged me to address this worry.

Explaining Supervenience

people of the sort we all know who also had the characteristics definitive of being an ideal observer would have the specified attitude of approval or disapproval towards it. And even in a world with no normal agents the counterfactual judgement could be true. In a world with only psychopaths some of the actions they do could be such that they would be approved of by (nonexistent in such worlds) normal humans who are fully informed, omniperceptive, and so on. Similarly, an action done in a ghost world can be such that a normal human who knew enough and so on would disapprove or approve of it. Variants of the ideal observer theory that substituted a dispositional analysis for a counterfactual analysis would be in a position to do the same thing. On such views the rightness or wrongness of an action would be identified with or constituted by the dispositions of actual normal observers under conditions of full information, and so on¹⁰. And actions can be such as to interact with those dispositions (when contemplated by such observers in the actual world) even in worlds where there are no normal human beings.

Ideal observer theorists are not the only naturalists who can make this sort of move. David Lewis (1987) proposed something like this when he floated a dispositional theory of value. Here the valuer need not be an observer; they need only imagine the putative values. And the values would not have their weights assigned by the people doing the imagining. The analysis tells us what is and isn't a value, but not how valuable the values are. The view leads to a worry about the modal status of the resulting conclusions.

. . . Psychology is contingent. Our dispositions to value things might have been otherwise than they actually are. We might have been disposed, under ideal conditions, to value seasickness and petty sleaze above all else. Does the dispositional theory imply that, had we been thus disposed, those things would have been values? That seems wrong.

No: we can take the reference to our dispositions as rigidified. Even speaking within the scope of a counterfactual supposition, the things that count as values are what we are *actually* disposed to value, not those we would have valued in the counterfactual situation. No

¹⁰So far as I can see there is no obstacle for a naturalist who commits to reductions as identities from endorsing this sort of theory. Because the reduction connects the property to be reduced to a relatively unified property, there is no worry about identity generated by multiple properties all of which are sufficient for the reduced property so that only a disjunctive property would be both necessary and sufficient for the property up for reduction. And given that the result is nondisjunctive, the reduction itself could capture a real natural similarity shared by all property instances.

worries – unless seasickness actually *is* a value, it still wouldn't be a value even if we would have been disposed to value it. (Lewis 1989, 133)

Lewis's concern here was to address contingency, not to defend naturalism from the threat of vicious but nonactual ghosts. And he wasn't completely happy with rigidification as an answer to that particular unease. But as a strategy for constructing a perfectly naturalistic reduction base for value it works quite well. In general any theory that analyzes normativity in terms of some relation to the psychologies of agents or observers might restrict the relevant class to agents or observers of the sort found in the actual world. If, normative properties aside, such a world satisfies naturalistic constraints on what there is in this world, relational properties constructed out of a naturalistic relation to those agents or observers will themselves be natural in a good sense. And this will remain so even if non-natural stuff in counterfactual scenarios could stand in that relation to those natural agents or observers.

Such theories can explain supervenience easily enough. On the theory that normal reactions of approval or disapproval to particular actions track their descriptive features in regular ways, it would be expected that people would, in general, react to identical actions with the identical attitudes. Or, to take into account the influence of mood and idiosyncratic but normal psychologies, at least they are disposed to react similarly even if in actual fact they don't always react in precisely the same way. This is a point in favor of dispositional rather than counterfactual developments of this general strategy. But once it is taken on board it is pretty easy to see how these theories would explain supervenience. So long as our dispositions to react tracked descriptive similarities and differences between actions, this way of analyzing normative properties would have the extension of these properties supervene on the extension of the properties they were dispositions to react to.

There are thus at least two viable naturalist strategies for explaining supervenience. Neither relies just on the fact that the supervenient normative property is in the same general metaphysical category as the properties in the subvenient supervenience base. Each relies on finding some level of description at which the normative property might be identified with a real natural similarity among the things in its extension and then offering a further story about why this property is determined by the presence or absence of distinct properties in the supervenience base. This second further component is essential to these theories being able to explain supervenience. And that is

reason to think that it isn't the naturalness of the normative property doing the explanatory work but rather something more specific.

6. Back to NonNaturalism

We should now briefly consider whether nonnaturalists are in a position to offer parallel explanations. While it is hard to be sure without a precise delineation of the boundary between naturalism and nonnaturalism, it looks to me as though most of the naturalist explanations of supervenience available can be mimicked by nonnaturalists. I'll work backwards.

There should be nonnaturalist theories that look a lot like actualist relationalism. John McDowell's (1985) secondary quality model for value is not obviously naturalistic. It can be very roughly captured by the idea that an object of evaluation is valuable iff it merits a certain sort of response from an appropriately sensitive and informed agent. Applied to right and wrong actions it might be something like, an action is right iff an appropriately sensitive and informed agent would be motivated to do it in the circumstances. An actualist version of the view could be constructed by adding a clause to ensure that the appropriately sensitive agent be like ourselves in other respects. Since merit seems itself to be a normative matter, and since his account does not reduce it to other nonnormative properties, a nonnaturalist could accept this account. Coupling this idea to a theory of merited action which makes merit a function of the nonnormative descriptive features of the actions up for assessment would offer an explanation of the supervenience of the value properties on the nonnormative descriptive properties.¹¹

I'm of two minds about whether this is really parallel to the explanation of supervenience offered by rigidified relational naturalism. That theory plugs in an empirical psychological theory to explain why the dispositions of ideal observers or agents will track the nonnormative descriptive differences of the actions up for assessment. The present nonnaturalist theory plugs in an a priori moral thesis – that merit itself supervenes on the nonnormative descriptive features of actions – to do the same work. The claim plugged in strikes me as true and not all that controversial. And in a way the a priori nature of the claim is an advantage, insofar as the truth of the supervenience constraint is supposed to be a priori available to us. On the other hand, the explanation is of a different sort than the naturalist's, relying on conceptual truths about the nature of merit rather than psychological laws to secure the

¹¹I think that a similar story could be told using Scanlon's (2014) irreducibly normative 4 place reason relation and some of the explanation would have to be changed to do it. But I omit that for reasons of space.

connection between the properties. The nonnaturalist's explanation is in this respect closer to R. M. Hare's (1952, pp. 130 ff.) explanation of supervenience (via a meaning rule in virtue of which moral judgements commend on the basis of their descriptive features) than the reductive naturalist's. One can worry that the principle regarding merit used is very close to the phenomenon that we are trying to explain and that we are thus moving in a circle.

Perhaps this unease should push the nonnaturalist towards moral functionalism. For a moral functionalist will advert to the complex role that specific moral properties play in our lives. Being right is action guiding in various ways. And the moral functionalist can argue that we must therefore have some grip on which things are right and which are not. This grip will have to be mediated by the other properties an action can have and that we have epistemic access to. So the role that rightness plays requires that we regularly be able to figure out the moral status of a prospective action based on our knowledge of its other features. For all that, such a functionalist can think that the action-guiding role of rightness is still irreducibly normative in virtue of containing some irreducibly normative aspects. They might think that it is part of the role of moral properties that they be apt to figure in justifications and they might think that this cannot be analyzed away. At the same time they can agree with the nonreductive naturalist that a given instance of a deontic moral property can be realized by some instance of a specific natural property so that it is right in virtue of having that complex natural property. The views will be quite parallel, except for the nonnaturalist's insistence that the role is irreducibly normative whereas the naturalist will think it is not.

This leaves me thinking that the explanatory challenge doesn't in general favor naturalism or nonnaturalism, though it may well count against particular versions of either one.¹²

Bibliography

- Armstrong, David. M. 1978. *Nominalism and Realism: Vol. 1: Universals and Scientific Realism*. Cambridge: Cambridge University Press.
- Bader, Ralph. 2017. "The Grounding Argument against Non-Reductive Moral Realism." In Shafer-Landau (ed.) *Oxford Studies in Metaethics* 12, 106–134. New York: Oxford University Press.

¹²Thanks to John Brunero, Jennifer McKittrick, Joe Mendola and Eric Wiland for discussion as I was writing this paper, and to Jamie Dreier and Tristram McPherson for discussion of related issues over the years. Thanks also to a helpful referee for this journal and to Jennifer Haley for her comments on an earlier draft.

Explaining Supervenience

- Bennett, Karen. 2017. *Making Things Up*. Oxford: Oxford University Press.
- Bhagal, Harjit. 2022. Moral Principle Explanations of Supervenience. *Philosophical Studies*: 1–20.
- Blackburn, Simon. 1984. *Spreading the Word*. Oxford: Clarendon.
- Dancy, Jonathan. 2005. “Nonnaturalism.” In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 91–121. Oxford: Oxford University Press.
- Dorsey, Jonathan Eric. 2016. “On the Grounding-Reduction Link.” *Australasian Journal of Philosophy*, 53, 411–422.
- Dreier, James. 1992. “The Supervenience Argument Against Moral Realism.” *The Southern Journal of Philosophy* 30: 13–38.
- Dreier, Jamie. 2019. “Is there a supervenience problem for robust moral realism?” *Philosophical Studies* 176 (6): 1391–1408.
- Dunaway, Billy and Tristram McPherson. 2016. “Reference Magnetism as a Solution to the Moral Twin Earth Problem.” *Ergo: An Open Access Journal of Philosophy* 3, no. 25: 639–679.
- Elliot, Aaron. 2014. “Can Moral Principles Explain Supervenience? *Res Philosophica*, Special Issue: Moral Nonnaturalism, Vol. 91, No. 4: 629–659.
- Faraci, David. 2017. “Hybrid Non-Naturalism Does Not Meet the Supervenience Challenge.” *Journal of Ethics and Social Philosophy* 12 (3).
- Fine, Kit. 2002. “The Varieties of Necessity.” In T. Gendler and J. Hawthorne, eds. *Conceivability and Possibility*. Oxford: Oxford University Press: 253–281.
- Fogal, Daniel and Olle Risberg, 2020. “The Metaphysics of Moral Explanations.” *Oxford Studies in Metaethics* 15, Oxford: Oxford University Press: 170–194.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. New York: Oxford University Press.
- Jackson, F. And Phillip Pettit. 1995. “Moral Functionalism and Moral Motivation,” *The Philosophical Quarterly* 45 (178): 20–40.
- Jackson, F. And Phillip Pettit. 1996. “Moral Functionalism, Supervenience and Reductionism.” *The Philosophical Quarterly* 46 (182): 82–86.
- Kramer, Matthew. 2009. *Moral Realism as a Moral Doctrine*, Oxford: Blackwell.
- Leary, Stephanie. 2017. “Non-Naturalism and Normative Necessities.” In *Oxford Studies in Metaethics* 12, edited by Russ Shafer-Landau, 76–105. New York: Oxford University Press.
- Lewis, David. 1983. “New Work for a Theory of Universals.” *Australasian*

- Journal of Philosophy* 61, no. 4: 343–377.
- McPherson, T. 2012. “Ethical Non-naturalism and the Metaphysics of Supervenience,” In *Oxford Studies in Metaethics* 7, edited by Russ Shafer-Landau, 205–34. Oxford: Oxford University Press.
- McPherson, T. 2015. “What is at Stake in Debates Among Normative Realists?” *Noûs* 49 (1): 123–146.
- McPherson, Tristram. 2021. “Metaphysical Structure for Moral Realists.” In *The Oxford Handbook of Moral Realism*, edited by Paul Bloomfield and David Copp. Oxford: Oxford University Press.
- Morton, Justin. 2020. “Grounding the normative: a problem for structured non-naturalism.” *Philosophical Studies* 177 (1): 173–196.
- Moore, George E. 1903. *Principia Ethica*. Cambridge UK: Cambridge University Press.
- Railton, Peter. 2017. “Naturalistic Realism in Metaethics.” In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 151–169. New York: Routledge.
- Rosen, Gideon. 2010. “Metaphysical Dependence: Grounding and Reduction,” in B. Hale and A. Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*, New York: Oxford University Press, 109–36.
- Rosen, Gideon. 2022. “What is Normative Necessity?” In *Metaphysics, Meaning and Modality: Themes from Kit Fine*, edited by M. Dumitru. Oxford: Oxford University Press. 205–233.
- Scanlon, T. M. 2014. *Being Realistic About Reasons*: Oxford: Oxford University Press.
- Schroeder, Mark. 2007. *Slaves of the Passions*. Oxford: Oxford University Press.
- Setiya, Kieran. 2012. *Knowing Right from Wrong*. Oxford: Oxford University Press.
- Sinhababu, Neil. 2018. “Ethical Reductionism,” *Journal of Ethics and Social Philosophy* 13 (1): 32–52.
- Sturgeon, N. 2006. “Ethical Naturalism,” in D. Copp (ed.) *The Oxford Handbook of Ethical Theory*, 91–121.
- Suikkanen, Jussi. 2010. “Non-Naturalism: The Jackson Challenge,” In *Oxford Studies in Metaethics* 5, edited by Russ Shafer-Landau, 87–110. New York: Oxford University Press.
- Toppinen, Teemu. 2018. “Essentially Grounded Non-Naturalism and Normative Supervenience.” *Topoi* 37 (4): 645–653.
- van Roojen, Mark. 1996. “Moral Functionalism and Moral Reductionism,” *The Philosophical Quarterly*, 46 (182): 77–81.
- van Roojen, Mark. 2006. “Knowing Enough To Disagree: A New Response

Explaining Supervenience

- to the Moral Twin Earth Argument,” In *Oxford Studies In Metaethics* 1, edited by Russ Shafer-Landau, Oxford: Oxford University Press: 161–193.
- van Roojen, Mark. 2015. *Metaethics: A Contemporary Introduction*. New York: Routledge.
- van Roojen, Mark. Forthcoming “Does Anything We Care about Distinguish the Non-Natural from the Natural?,” in Paul Bloomfield and David Copp (eds.) *Oxford Handbook of Moral Realism*. Oxford: Oxford University Press.
- Väyrynen, Pekka. 2017. “The Supervenience Challenge to Non-Naturalism.” In Tristram McPherson & David Plunkett (eds.), *The Routledge Handbook of Metaethics*. New York: Routledge: 170–84.

Mark van Roojen
University of Nebraska–Lincoln
mvrl@earthlink.net