



Arguing About Moral Causation

Umut Baysan¹

Accepted: 29 May 2024
© The Author(s) 2024

1 The Argument from Moral Causation

Within the span of two decades, hundreds of Post Office local branch managers (also known as “subpostmasters” in the UK) were fined, prosecuted, and convicted for allegations of theft, fraud, and false accounting. Some subpostmasters served prison sentences, many others had to pay very large amounts of money to the Post Office. We now know that the accusations which led to these fines, prosecutions, and convictions were baseless, as they were caused by systematic failures of the accounting software that the Post Office was using at the time. Although the faults of the accounting software were reported again and again, the senior management at the Post Office chose to ignore such reports and instead went ahead with prosecutions. While it took two decades for the truth to come out, over the years, hundreds of subpostmasters lost their businesses, homes, and families. Many of them had to live with criminal records, and some of them died before their convictions could be overturned. The Post Office’s treatment of the subpostmasters is a clear example of *injustice*. Since these events have been popularised thanks to a TV series and the subsequent media coverage, there has been a public outrage in response to Post Office’s treatment of the subpostmasters. The government had to take an explicit stance on what we now call the “Post Office scandal”, and the Prime Minister announced plans to overturn outstanding convictions and facilitate plans to compensate the victims swiftly.¹

Based on these events, we might want to say that the Post Office’s unjust treatment of the subpostmasters caused public outrage. On the face of it, this would be to make a *causal* statement that attributes causal powers to a *moral* property: the property of *being unjust*. As such, this would be one of many examples of statements that appear to give moral properties roles in causation. Proponents of *naturalism* in metaethics typically think that the truth of claims of this sort supports their view. Sturgeon, for one, is explicit about this:

¹ See BBC News (2024).

✉ Umut Baysan
umut.baysan@philosophy.ox.ac.uk

¹ University of Oxford, St Anne’s College, 56 Woodstock Road, Oxford OX2 6HS, UK

The most compelling ground for ethical naturalism, in my view, lies in the fact that goodness and other ethical properties appear to play a causal role in the natural world, and that it is plausible to think that any property that can play such a role must be a natural property. (2003, 538; see also Sturgeon 2005 for very similar remarks).

Call this *the argument from moral causation*. While there are historical precursors to this argument² as well as more recent revivals of this line of thinking³, in what follows, I will work with the formulation found in the quote from Sturgeon above, and present it in terms of two central premises:

- (P1) Moral properties play causal roles in the natural world.
- (P2) Any property that plays a causal role in the natural world is a natural property.
- (C) Therefore, moral properties are natural properties.

We should distinguish the argument from moral causation from another argument, namely one which concludes that moral properties *exist* because they play causal roles in the natural world. Call this latter argument *the causal argument for realism*. Showing that moral properties have causal efficacy shows, at the very least, that they exist. So, considerations that seem to support the argument from moral causation seem to also work towards meeting a challenge for realists, famously set by Harman (1977). This is the challenge that moral properties are not causally or explanatorily indispensable: it seems in principle possible to causally explain the occurrence of public outrage that followed the Post Office scandal purely by citing the non-moral properties of the preceding events. If this is true, it seems to support a form of anti-realism, at least insofar as we accept a causal/explanatory criterion of existence, which says that a property must be causally/explanatorily indispensable in order to earn its place in our ontology. Now, while the argument from moral causation and the causal argument for realism are continuous (especially for naturalists), it is important to treat them as different arguments for the simple reason that it should remain an open question that a property may play a role in the natural world without thereby being a natural property---a possibility I shall consider below.⁴ So, my primary focus in what follows will be the argument from moral causation (and not the causal argument for realism).

As simple and elegant as it is, the argument from moral cannot be maintained--or so shall I argue. After explaining the idea behind P1 (§2), I will argue that, based on the existing dialectic, the most likely way of defending P2 relies on certain assumptions about causation (§3), but these assumptions make P1 very implausible (§4). I will then consider ways of revising the argument, but it will transpire that the most plausible version of this argument cannot be a successful defence of naturalism

² See Sturgeon's earlier work (e.g., 1985), as well as Railton (1985), Boyd (1988), and Brink (1989).

³ See Majors (2003), Luco (2019), and Stringer (2019; 2021).

⁴ This should remain an open question because it is a possibility that causal indispensability and explanatory indispensability are separable (see Baron et al. 2021) or that P2 is false. I discuss these in §5 and §6 below.

(§5). I will conclude by considering a way of making sense of realism about moral properties without thereby giving them causal roles in the natural world (§6).

2 Moral Causation

P1 states that moral properties play causal roles in the natural world. I will take this to mean that there is moral causation. For there to be moral causation in this sense, moral properties (or instantiations thereof) need not be *sufficient* causes of any events that take place in the natural world. First, moral properties (or instantiations thereof), can be effects (rather than causes) in causal relations. Second, even if moral properties may figure as causes (not just effects), it may be that a moral property is only *a* contributing cause (rather than *the* cause) of a given event. Or it may be that a moral property (or instantiation thereof) is merely a part of background conditions in a causal relation. As far as the causal argument is concerned, the obtaining of any of these conditions will be sufficient for there to be moral causation. That said, in what follows, I will consider cases where moral properties are stipulated to figure as causes (so I will set aside cases where they are merely effects or background conditions), but I will leave it open that they may or may not be sufficient causes.

Suppose you see a group of children who are setting a cat on fire. You see what the children are doing and come to believe that what they are doing is wrong. Let's assume that what the children are doing is wrong. Here, we can make a case for moral causation if we are prepared to accept that the *wrongness* of the children's treatment of the cat causes you to judge that what they are doing is wrong.⁵ Following Stringer (2021, 347), we can call cases like this "judgement cases". These are called as such because the *explananda* in these putative explanations are judgements. Our example concerning the Post Office scandal should also be classified as a judgement case: presumably, if the injustice that the subpostmasters faced plays any causal role in bringing about public outrage, it does so via causally contributing to the individual judgements of those who are outraged by what has happened.

Stringer helpfully contrasts judgement cases with what he calls "compensation cases" (2021, 349). It is worth examining these cases as they are metaphysically interesting, and to the best of my knowledge, they have not received much attention in discussions of moral causation. As Stringer puts it, these are "cases where

⁵ This example is from Harman (1977), who uses it to draw a different conclusion. For Harman, intuitively, this action's instantiation of *being wrong* is causally redundant with respect to your coming to believe that what the children are doing is wrong; insofar as you accept the general moral claim that it is wrong to set a cat on fire, you are in a position to judge that what the children are doing is wrong. Now, suppose that your judgement is false; perhaps the principle in question is false because there is no such thing as right or wrong; or perhaps it is false because, for whatever reason, it not wrong to set a cat on fire; or perhaps, children cannot bear moral responsibility; or perhaps the children are setting the cat on fire to rescue five other cats and, as it happens, a moral theory which condones setting one cat on fire to rescue five other cats is correct. The fact that the children's action instantiates the property of *being wrong* is not required for causally explaining the fact that you come to believe that what the children are doing is wrong because insofar as you accept the principle in question (even if it is false), you are in a position to form the same moral judgement.

we wrong others and thereby owe them compensation, [so] ...the wrongness of our actions has effects on the moral statuses of future actions or omissions” (ibid.). Assuming that the Post Office’s treatment of the subpostmasters was wrong, it is plausible to think that the unjustly treated subpostmasters are owed compensation. Now, suppose that the Post Office compensates the subpostmasters. This subsequent action of compensating the victims has the property of *being obligatory*. The fact that this subsequent action is obligatory is explained by the fact the Post Office’s previous treatment of the subpostmasters was wrong. If there were no wrong treatment in the first place, the subsequent compensation would not have been obligatory. Perhaps, it would have been supererogatory---or even morally impermissible depending on what further costs such compensation entails. Stringer acknowledges that cases like this may look counterintuitive as cases of causation, but he correctly notes that these are cases in which moral properties function as *difference-makers*: “the presence of wrongness makes a difference because without the previous wrongdoings, the same [subsequent] actions do not have the same moral statuses they have” (ibid.). If we assume that difference-making is a good candidate as a distinctive feature of causation, we can follow Stringer in thinking these are cases of moral causation.

3 From Causal to Natural?

Suppose we agree that the judgement cases and the compensation cases we have seen are indeed cases of moral causation, so accept that moral properties play causal roles in the natural world P1. According to the naturalist line of thinking that I shall scrutinise below, the existence of moral causation very strongly supports naturalism: moral causation is not only sufficient for moral properties to exist, but it is also “the most compelling ground” for naturalism (Sturgeon 2003, 538). As I have noted in §1, this rests on P2, i.e., the claim that any property that plays a causal role in the natural world is a natural property.

What might be the reason to accept P2? Now, there is a conception of natural properties according to which, by definition, a property is natural just in case it is capable of playing causal roles in the natural world. On this conception of a natural property, P2 would of course be true. But what is a reason to accept this conception independently of deeming it true by definition? In Sturgeon’s own writings, the case for P2 rests on considerations that have to do with the causal *completeness* of the natural world:

[A] philosophical naturalist will believe that the mere fact that a property plays a causal role in the natural world provides a good reason for thinking that it is itself a natural property. I do not mean that we know a priori that only natural properties could have natural effects. For all we can know a priori, there might be gods or angels who produce natural effects in virtue of their specifically supernatural properties ... But philosophical naturalists do not believe in supernatural entities of these sorts and believe, indeed, that *they have powerful*

philosophical arguments against the existence of any of them. (2005, 101-102, emphasis added)

But what are these “powerful philosophical arguments” against the causal efficacy of entities that are not natural? Sturgeon’s remarks in a footnote appending these claims give an answer: “The [aforementioned] argument ... was suggested to me by reflection on a standard argument for physicalism about the mental from the causal efficacy of the mental and the *completeness* of physics” (ibid., n. 13, emphasis added).⁶ Based on this, it is reasonable to think that the argument for P2 should go as follows.

Take some property M that can be a cause of some event in the natural world. Let’s assume that an instance of M, at a given time, is a cause of some event e in the natural world. Let’s call this assumption (M-Causation):

(M-Causation) M causes e.

How do we get the conclusion that M must be a natural property given that an instance of it is a cause of a natural event? First, as suggested by Sturgeon, we can assume a completeness claim concerning the natural domain:

(N-Completeness) If an event in the natural world has a cause at a time, it has a sufficient *natural* cause at that time.⁷

Note that (M-Causation) and (N-Completeness) do not entail that M must be a natural property. All they entail is that e has a sufficient natural cause at the time at which M causes it. (Let’s say that that time is t_1 .) Let’s take P to be a natural property whose instantiation at t_1 is a sufficient cause of e. This picture is entirely consistent with the claim that M is not identical with P (or any other natural property), as e may have more than one cause (at least one of which is a sufficient cause) at t_1 . The next step in the argument is to state the following:

(Exclusion) No event has more than one simultaneous cause (where at least one such cause is a sufficient cause) unless it is causally overdetermined.

In other words, except for cases of causal overdetermination, a given sufficient cause for an event will exclude any other putative simultaneous cause. Again, note that (M-Causation), (N-Completeness), and (Exclusion) together do not entail that M is identical with P (or any natural property). How do we block this consequence? Of course, by stipulating that this is not a case of causal overdetermination. A non-*ad hoc* way of doing this would be to put a ban on systematic causal overdetermination more generally:

⁶ Nearly all parties to the debate on moral causation draw parallels with what is known as “the causal exclusion argument” in philosophy of mind popularised by Kim (1998; 2005), which does rely on claims about the causal completeness of the physical and certain theses about causation (see Shafer-Landau 2003; Majors 2003; Cuneo 2006; Stringer 2021; Baron et al. 2021).

⁷ In the quoted passage above, Sturgeon makes a reference to the causal completeness of the *physical* (P-Completeness), rather than the *natural* (N-Completeness). While it is important for my argument in §4 that Sturgeon accepts the causal completeness of the physical, the argument for P2 requires the causal completeness of the natural.

(No-Overdetermination) There may be rare occasions of causal overdetermination, but there is no systematic causal overdetermination in the natural world.

This is a non-*ad hoc* way of getting the desired consequence, as it does not just say that M and P do not causally overdetermine e. After all, this is not a thesis specifically about M, P, and e. But it suggests that M and P do not causally overdetermine e, as in all relevant cases where we are interested in some property that “M” stands for, if causation by M were a case of causal overdetermination, there would have been systematic causal overdetermination. In cases where “M” stands for mental properties, causal overdetermination by M entails that *whenever* there is mental causation, there is causal overdetermination. In cases where it stands for moral properties, causal overdetermination by M entails that whenever there is moral causation, there is causal overdetermination. (No-Overdetermination) is posited to rule out systematic causal overdetermination of this sort without ruling out the possibility of rare occasions of causal overdetermination, such as the “firing squad” cases in which multiple shooters simultaneously shoot a victim and where each individual shot is sufficient to kill the victim.

We now have the required premises to get the conclusion that any property that can play a causal role in the natural world is a natural property. The most natural way these four claims can be true all at once is through an identification of M with P. If M is indeed the same property as P, then it can be true that M is a cause of e at t_1 (M-Causation), e has a sufficient natural cause at t_1 (N-Completeness), namely M, which is identical with P. All this is compatible with both (Exclusion) and (No-Overdetermination). Causation of e by M at t_1 and causation of e by P at t_1 after all are not different causal relations. Hence, we have a causal exclusion argument suggesting that the only way for M not be excluded from causing e is via its identification with P.

Why should we believe any of this? We are merely assuming (M-Causation) to see what it implies. More substantially, (N-Completeness) is a corollary of what is known as “the causal closure of the physical”, which states:

(P-Completeness) If a physical event has a cause at a time, it has a sufficient physical cause at that time.

Proponents of (P-Completeness) often take this to be an empirically motivated claim about physics with very strong inductive support. Of course, (P-Completeness) itself is not a principle of physics; moreover, it is not entailed by any law of physics, and its proponents acknowledge that this is not a principle that one can find in a physics textbook (Papineau 2001, 14). That said, a widely shared view is that rejecting this principle is comparable to “betting against the truth of physics” (Lewis 1988, 95). According to Papineau, this inductive support comes from the fact that the history of physical sciences is full of attempts to find non-physical causes of physical phenomena, yet such attempts have failed at each time. Papineau appeals to an example from physiology: if (P-Completeness) were false, we would have discovered “special”, *sui generis*, irreducibly non-physical forces that move our bodies. “If there were such forces, they could be expected to display some manifestation of their presence. But detailed physiological investigation failed to uncover evidence of anything except

familiar physical forces” (ibid., 31). A related argument is from McLaughlin (1992). McLaughlin’s discussion concerns the views and arguments of British emergentists of the early 20th century (e.g., Broad 1925). McLaughlin suggests that “many of the central doctrines of [British] emergentism fly in the face of many of the major scientific achievements of [the 20th] century” (1992: 39). Given that a central commitment of British emergentists was the postulation of special forces that violate (P-Completeness), the fact that emergentism of this sort was abandoned due to scientific developments of the century, suggests, according to McLaughlin, that physical sciences provide indirect support for (P-Completeness). I should add to these the further observation that Sturgeon, who explicitly states P2 as part of the argument from moral causation, makes a reference to (P-Completeness), citing Papineau’s (2001) aforementioned argument favourably (Sturgeon 2005, 102, fn. 13).

What about (Exclusion) and (No-Overdetermination)? Kim, who is one of the most prominent proponents of the causal exclusion argument in philosophy of mind, states that (Exclusion) is “virtually an analytic truth” (2005, 51). My view is that we should think of (Exclusion) not as an analytic truth, but as a general metaphysical thesis about causation, and whether we should accept it or not depends on how we should understand causation more generally.⁸ Unlike (Exclusion), no one takes (No-Overdetermination) to be a “virtually analytic truth”. I think we should think of (No-Overdetermination) as a principle whose status is very similar to a principle of ontological parsimony. My reading of this principle along these lines is supported by one of the very few explicit arguments for it, which is due to Schiffer (1987). Considering the possibility of systematic causal overdetermination in the case of mental causation, Schiffer argues as follows:

If this sort of causal overdetermination obtained, then a mental event could *never* cause a bodily movement *except* in a case of causal overdetermination where there was a simultaneous and distinct neural cause of the movement ... This causal superfluosity is hard to believe in; it is hard to believe that God is such a bad engineer. (ibid., 148, emphasis added)

This concludes my presentation of the causal exclusion argument that supports P2. The argument is that for any event in the natural world that is caused by some property M, there will be a sufficient natural cause of that event, and unless M is itself a natural property, causation by M will imply a problematic sort of causal overdetermination. In the next section, I will argue that the reasoning behind this argument generates a problem for naturalists’ own treatment of moral causation.

⁸ See Baysan (2022) for a more detailed treatment of the idea that whether we should accept this principle depends on how we understand causation more generally.

4 The Causal Argument Undermined

Given the arguments of §3, we can justifiably speculate that the best case for the claim that only natural properties can play causal roles in the natural world requires accepting (Exclusion) and (No-Overdetermination). Henceforth, I will call the conjunction of these two principles *exclusionism about causation*. In this section, I want to show that exclusionism about causation leads to a difficulty for accounting for moral causation on naturalists' own terms.

To show that the difficulty that naturalism faces, I will make a distinction between two varieties of naturalism: *reductive* and *non-reductive*.⁹ There is no consensus on how we should use these terms, so let me explicitly state what I mean. According to my construal of naturalism, both varieties of naturalism are committed to the view that moral properties *are* natural properties. So, for any moral property M, naturalism is committed to the claim that there is some natural property N such that $M=N$. But this is not to be confused with reductive naturalism, at least according to how I understand this view. Reductive naturalists have the further commitment that there is some natural property N^* such that N^* is either a paradigmatic natural property or a disjunction of paradigmatic natural properties, and moreover $N=N^*$ (hence $M=N^*$). Perhaps a form hedonism is true and the property of *being good* is identical with the property of *being pleasant*, which is, let's suppose, in the purview of neuroscience. Alternatively, we can follow Jackson (1998) in thinking that N^* is the disjunctive property that comprises all individual properties (as disjuncts) that (can) occupy the role of *being good* according to an idealised folk theory of morality and take *being good* to be identical with that disjunctive property. Thus, according to my construal, these views are examples of reductive naturalism.

Non-reductive naturalists reject the identity claim regarding M and N^* . They might reject the hedonist view I mentioned above on the basis that *being pleasant* is only one of many different ways of *being good*. And they might reject views like Jackson's because they think that moral properties are not disjunctive properties. (Perhaps, they think that there are no disjunctive properties *simpliciter*, so it is not surprising that moral properties are not disjunctive properties.) Either way, they reject the identity of M and N^* because either N^* is one of many different ways of being or contributing to M or there is no suitable candidate for N^* that can be identified with M. Importantly for the arguments of this paper, non-reductive naturalism is still committed to the identity of moral properties with natural properties. As Sturgeon puts it, "if [good] is a natural property it is also identical with a natural property ... because it is, unsurprisingly, identical with itself" (Sturgeon 2005, 99; for other defences of non-reductive naturalism along similar lines, see Sturgeon 1985; Boyd 1988; Brink 1989).

⁹ I understand reduction (in the relevant sense) to be an ontological relation: A reduces to B only if A is identical to B. To be clear, there is another sense of "reduction" on which A reduces to B only if A is definable in terms of B. Although there are connections between reductive (and non-reductive) theories in the former, ontological, sense and reductive (and non-reductive) theories in the latter, definitional, sense, for reasons of space, I will set this issue aside.

Let me illustrate the problem for the causal argument by focusing on non-reductive naturalism first. If exclusionism about causation is true, a moral property *M* cannot be causally efficacious in the natural world if there are *physical* properties that are causally sufficient for the effects we associate with *M*. So, if exclusionism about causation is correct, then moral properties that are identical with natural properties can be causally efficacious only if we *reject* the claim that the effects we associate with moral properties have sufficient physical causes. But that would amount to rejecting (P-Completeness). So, if naturalists want to say that moral properties can be causally efficacious in the natural world, and at the same time hold exclusionism about causation, they would have to be *committed* to the failure of (P-Completeness). Now, while naturalism is not necessarily committed to (P-Completeness), I take it that it would be a significant cost to the overall plausibility of naturalism if entails the violation of (P-Completeness). Moreover, on a somewhat *ad hominem* note, while naturalism as a view is not committed to (P-Completeness), naturalist philosophers typically accept it. Sturgeon (2005) accepts it as part of his philosophical naturalism more broadly; so does Jackson (1998), whose variant of naturalism is an instance of physicalism more generally---and physicalism, as a view, is committed to (P-Completeness).¹⁰

Of course, naturalists might want to avoid this problem by *identifying* moral properties with physical properties. While I concede that some proponents of naturalism are open to this possibility (e.g., Boyd 1988, 199)¹¹, I simply assume that this is very implausible, and it is a significant cost to the overall plausibility of naturalism if it implies that moral properties are physical properties. While it is not quite an argument, this assumption is backed by the intuition that moral properties are “just too different” from physical properties to be identical to them.¹²

Is this not a problem for *all* forms of non-reductive views in philosophy of science? That is, can we not say the same thing about psychology, biology, chemistry and so on?¹³ Yes, and no. “Yes”, because, this is indeed one of Kim’s (1998) main aims in his defence of the causal exclusion argument. According to Kim, given exclusionism about causation and the presumed truth of (P-Completeness), mental properties and the properties that the special sciences study are either causally inefficacious or otherwise identical with physical properties. So, if (P-Completeness) is true, exclusionism about causation implies that only physical properties can be causally efficacious in the physical world. The fact that this rules out moral properties playing any causal roles in the physical world---assuming that moral properties are not physical properties---is just an extension of this more general idea. But, also,

¹⁰ See Wilson (1999) for a compelling argument that (P-Completeness) is non-negotiable for physicalists.

¹¹ I am grateful to an anonymous reviewer for pointing this out.

¹² This is obviously analogous to the *just-too-different* objection to naturalism (Enoch 2011). However, it is importantly more modest than the more traditional just-too-different objection, as I am contrasting the moral with the physical, rather than with the natural. That is, I am granting to the naturalist that moral properties may be natural properties; I am simply rejecting their identification with physical properties on intuitive grounds.

¹³ See Majors (2003) and Baron et al. (2021) for related discussion.

“No”, because proponents of non-reductive views in philosophy of science are in a position to reject exclusionism about causation, precisely because of these worries. But how can they do this?

A case against exclusionism about causation may be made by acknowledging the plausible claim that when there is a necessary connection between two putative simultaneous causes of a subsequent event, they are not distinct enough to generate genuine causal overdetermination. Although there are various ways this idea can be spelled out, Bennett’s (2003) diagnosis of this issue is helpful for my purposes here, thus I will work with this diagnosis.¹⁴ As Bennett correctly observes, for two events A and B to causally overdetermine E, it has to be non-vacuously true that:

(OD-1) If A had not happened but B had happened, then E would still have happened; and

(OD-2) If B had not happened but A had happened, then E would still have happened.

This is how we understand causal overdetermination in firing squad cases. To illustrate: Amy and Barny simultaneously shoot Casey. Each shot is lethal. If Amy had not fired, Barny would still have, and Casey would still have died. If Barny had not fired, Amy would still have fired, and Casey would still have died. Now, contrast these with cases where A and B are not metaphysically distinct. If there is a metaphysically necessary connection between A and B such that it is not metaphysically possible for A to occur without B, (OD-2) is only vacuously true.

The lesson from this is that when there are such necessary connections between putatively causally overdetermining causes of an event, we end up with counterexamples to exclusionism about causation. In cases of mental causation, a non-reductive physicalist can say that my pain experience and the physical properties that realize that experience are individually sufficient simultaneous causes of my behaviour (e.g., of saying “ouch”), but they do not causally overdetermine it. After all, it is not metaphysically possible for the physical realizers of my pain experience to be instantiated without me having a pain experience. Now, if proponents of non-reductive views in inter-level philosophy of science can reject exclusionism about causation, so should non-reductive naturalists in ethics.¹⁵ And here is a very good reason for them to do so. The key is to acknowledge:

¹⁴ For other ways of rejecting this principle, see, among others, Wilson (2011), Árnadóttir & Crane (2013) and Baysan (2018).

¹⁵ See Baron et al. (2021) for related discussion. Baron et al. draw slightly different conclusions from this, as they maintain that statements like (OD-2) can be explanatory even when there is a necessary connection between A and B (such that it is impossible for A to obtain without B obtaining). Such statements are known as counterpossibles (counterfactuals with impossible antecedents) and they have explanatory power in mathematics, they argue. (Consider: “If 7 weren’t a prime number ...”). Now, if counterpossibles in other domains have explanatory power, it is possible that they do have explanatory power in ethics too, and this might just be the solution to the exclusion problem applied to moral properties. However, it should be noted that Baron et al. argue, convincingly in my view, that explanatory power does not always amount to *causal* power. For example, when counterpossibles in mathematics have explanatory power, what they explain need not be underwritten by causal relations. I will come back to this point in §6.

(Moral Supervenience) It is impossible for two states of affairs to be identical in all natural properties but different in their moral properties.

Imagine two cat-burning events that instantiate the same natural properties. Naturally indiscernible cats are set on fire in indiscernible ways by naturally indiscernible children, in naturally indiscernible circumstances. (Moral Supervenience) implies that, necessarily, if one of these cat-burning events involves a wrong action, so does the other. Likewise, if one does not involve any wrongdoing, neither does the other. This seems to suggest that the natural properties of an action metaphysically necessitate its moral properties.

Let's assume that the children's setting the cat on fire is indeed wrong, so it has the moral property M. Let's take N^+ to be the collection of natural properties that constitute the children's setting the cat on fire. If (Moral Supervenience) holds, then N^+ metaphysically necessitates M. The exclusion worry based on this case purports to show that any causal relevance we might want to attribute to M will be excluded by the causal relevance that N^+ has. This is because assigning any causal power to M implies accepting causal overdetermination, which would generalise to all putative cases of moral causation. This implies systematic causal overdetermination, which is ruled out by exclusionism about causation.

However, if (Moral Supervenience) holds, this problem has a straightforward solution. As we have seen in the previous section, if N^+ necessitates M, then we cannot have a situation in which an event instantiates N^+ without thereby instantiating M, in which case the non-vacuous truth of the following conditional does not hold:

(OD-3) If M had not been instantiated but N^+ had been instantiated, then j would still have happened.

Given (Moral Supervenience), (OD-3) is at best *vacuously* true. So, M and N^+ do not causally overdetermine j, meaning that the exclusion worry based on judgement cases can be met.

Is (Moral Supervenience) true? To be perfectly honest, I don't know.¹⁶ But it is important to note that *if non-reductive naturalism is true*, then (Moral Supervenience) is true too. After all, if moral properties are natural properties (as per naturalism), then no two states of affairs can be different in moral properties without being different in natural properties.

Let me now briefly show that *reductive* naturalists face the exact same problem insofar as they accept exclusionism about causation. Recall that, on the reductive version of naturalism, we identify M with N^* . Although this may look like a promising way of solving the problem of moral causation that is raised by exclusionism about causation, the prospects of this solution are as good as the view that N^* is a basic physical property. But note that N^* is not a basic physical property---even according to reductive naturalism. A reductive naturalist may think that *being good* is identical with *being pleasant*; or alternatively, she may think that *being good* is identical with a disjunction of first-order natural properties such as *being pleasant*.

¹⁶ I raise doubts about it in an unpublished manuscript. See also Rosen (2020).

But in either option, N^* is still a higher-level property that is not identical with a basic physical property. So, the same problem that I have illustrated about N arises for N^* .

Therefore, if (Moral Supervenience) holds, exclusionism about causation is false. We have also seen that if naturalism is true, then (Moral Supervenience) holds. This suggests that naturalists should reject exclusionism about causation. However, if naturalists should reject exclusionism about causation, then they cannot defend their view by the argument from moral causation. After all, exclusionism about causation is required for defending P2 of the argument from moral causation. This concludes my argument that there is a tension between the two premises of the argument from moral causation.

5 Moral Explanation without Moral Causation¹⁷

In the previous section, I argued that if (Moral Supervenience) holds, naturalists can solve the causal exclusion problem that their view faces. The causal exclusion problem for moral causation is that the non-moral natural properties of a state of affairs, e.g., its fundamental physical properties, seem to causally exclude its moral properties. The solution that I proposed requires rejecting exclusionism about causation, and the most straightforward way of doing this in the context of moral causation is via accepting (Moral Supervenience). However, showing that exclusionism about causation is false is one thing, showing that there is moral causation is another. More specifically, while successfully solving the causal exclusion problem for moral causation may be a *necessary* condition for showing that there is moral causation, it is not clear that it is a *sufficient* condition for it. For one thing, we might want to leave room for moral epiphenomenalism¹⁸, according to which moral properties supervene on natural properties, but they are causally inert. Relatedly, (Moral Supervenience) is a very broad thesis, often endorsed by non-naturalists too¹⁹, but non-naturalists reject the claim that moral properties are causally efficacious.²⁰ So, we should conclude that while (Moral Supervenience) is consistent with moral causation, it is not evidence of moral causation.

Consider again the compensation cases we have seen in §2. In these cases, a moral property of an action functions as a difference-maker for the moral properties of a subsequent action. For example, the wrongness of the Post Office's treatment of the subpostmasters makes a difference to the moral status of the Post Office's

¹⁷ I would like to thank an anonymous reviewer for changing my mind about the content of this entire section.

¹⁸ I thank an anonymous reviewer for raising this point.

¹⁹ See Shafer-Landau (2003, 95-95), Enoch (2011, Ch. 6.2.2.1), and others. For many non-naturalists, the question is not whether the moral supervenes on the natural; the question is whether such supervenience is explainable (see McPherson 2012 for discussion).

²⁰ See Shafer-Landau (2003, 60), Enoch (2011), Parfit (2011, vol. 2, 493). See also Hayward (2019) and Baron et al. (2021) for discussion.

subsequent actions. But should we understand the kind of difference-making in question as a form of causation?

Here is a reason why we should not take the difference-making relations in question as evidence of a causal relation between moral properties. If the difference-making in question is evidence of causation, then non-naturalists should also be committed to there being moral causation in these cases. Assuming that the example is unproblematic from a first-order normative ethical point of view, what explains the difference-making relation between wrongness and obligatoriness in this example is a first-order normative ethical principle that specifies the conditions under which a kind of action is obligatory. For example, this may be the principle that if one party wrongfully harms another party, then the former ought to compensate the latter for the harm that is brought about. This principle, if true, is sufficient to explain the difference-making relation in our example. By and large, such resources of first-order normative ethical thinking are available to non-naturalists in the same way they are available to naturalists. But given that non-naturalists reject moral causation, we shouldn't think of the difference-making relation in compensation cases as evidence of moral causation.

If neither (Moral Supervenience) in judgement cases nor the difference-making relation in compensation cases is evidence for moral causation, does that mean moral properties do not play any role in these cases? One interesting middle ground is that both (Moral Supervenience) and the kind of difference-making relation we have in compensation cases indicate that there is role for moral properties in *explanation*, but such explanation need not be understood as *causal* explanation.²¹ This is a promising option because showing that moral properties are explanatorily relevant can be a sufficient way of responding to the challenge posed by anti-realists (discussed in §1), and if we can separate the question of explanatory relevance from that of causal efficacy, a moral realist can defend her position without having to tackle the problem of moral causation.

If true, what does this tell us about the argument from moral causation? Recall that the argument was as follows:

- (P1) Moral properties play causal roles in the natural world.
- (P2) Any property that plays a causal role in the natural world is a natural property.
- (C) Therefore, moral properties are natural properties.

Given the problems I have raised, the argument clearly needs a revision. Now, if the phenomena we appealed to in order to make sense of P1 are the phenomena of (Moral Supervenience) and difference-making (of the relevant kind), then there is reason to think that P1 should be replaced with the following:

- (P1*) Moral properties play explanatory roles in the natural world.

²¹ This option is explored in detail in Baron et al.'s (2021) recent article on moral explanations.

Now, if we want the revised argument to be valid, P2 should also be replaced with the following:

(P2*) Any property that plays an explanatory role in the natural world is a natural property.

Admittedly, the revised version of this argument should be called *the argument from moral explanation*. While the argument from moral explanation can validly infer that moral properties are natural properties, there is reason to think that such an argument cannot be sound. While we can (and should) accept P1*, there is little reason to accept P2*. In fact, there is a strong reason to reject it: the relevant explanations are compatible with non-naturalism, and if non-naturalism is true, then moral properties are not natural properties. To put it differently, moral properties need not be natural properties to play explanatory roles. The most plausible revision of the argument should thus go as follows:

(P1*) Moral properties play explanatory roles in the natural world.

(P2**) Any property that plays an explanatory role in the natural world is a real property.

(C*) Therefore, moral properties are real properties.

While this revised version of the argument is an argument I would be very happy to accept, it is no more an argument for naturalism; it is an argument for moral realism, and moral realism come in both naturalist and non-naturalist forms.

6 Concluding remarks

We have seen that defending naturalism by appealing to the phenomenon of moral causation is not an easy task. First, the argument would have to make a convincing case that only natural properties can play causal roles in the natural world, and the most promising way of making this case generates difficulties for showing that there is moral causation in the first place. Second, it is questionable that the phenomenon that we have called “moral causation” is really a causal phenomenon, as there are reasons to think that the relevant cases that are supposed to support moral causation merely show that moral properties are explanatorily relevant, and a property can be explanatorily relevant without being causally efficacious. Such explanatory relevance, I have argued, is compatible with any form of realism, and as such, does not support naturalism over competing accounts.

In concluding, I want to address one final question: *If moral properties do not play causal roles, what kind of explanatory role might they play?* Note that this is a conditional question, and I am not saying that moral properties do not play causal roles. But I think we should be open to the possibility that moral properties play roles without playing *causal* roles. One promising idea here is that moral properties play *normative* roles by offering normative explanations, and it is not obvious that normative explanations *must* be causal explanations.

Consider again the public outrage in response to the Post Office scandal. Arguably, outrage is a fitting kind of reaction to what the subpostmasters experienced, and this is explained by the fact that what the subpostmasters experienced was unjust. Here, the property of *being unjust* plays a normative role in explaining the fittingness of outrage. If the subpostmasters had not been subjected to an unjust treatment, it would not have been fitting to feel outrage. Likewise, consider the fact that there is now a moral obligation to compensate the subpostmasters. Again, there is an explanation of this fact in terms of the wrongness of what the Post Office did to the subpostmasters, and this is a normative explanation: the wrongness of the Post Office's treatment of the subpostmasters normatively explains its moral obligation to compensate. These considerations support the claim that moral properties figure in normative explanations, and insofar as we have reason to accept the content of these explanations, we have reason to accept the existence of moral properties.²²

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Árnadóttir, Steinvor, and Tim Crane. 2013. There is No Exclusion Problem. In *Mental Causation and Ontology*, ed. S.C. Gibb, et al., 248–266. Oxford: Oxford University Press.
- Baron, Samuel, Mark Colyvan, Kristie Miller, and Michael Rubin. 2021. Non-naturalistic moral explanation. *Synthese* 198: 4273–4294.
- Baysan, Umut. 2018. Epiphenomenal Properties. *Australasian Journal of Philosophy* 96: 419–431.
- Baysan, Umut. 2022. The Pursuit of Neutrality in the Metaphysics of Emergence. *Analysis* 82: 159–169.
- BBC News. 2024. Post Office Scandal Explained: What the Horizon Saga is All About. Retrieved March 22, 2024, from <https://www.bbc.co.uk/news/business-56718036>.
- Bennett, Karen. 2003. Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract it. *Noûs* 37: 471–497.
- Boyd, Richard. 1988. How to be a Moral Realist. In *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord, 181–228. Ithaca, NY: Cornell University Press.
- Brink, David. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Broad, C.D. 1925. *The Mind and its Place in Nature*, 1925. London: Routledge & Kegan Paul.
- Cuneo, Terence. 2006. Moral Facts as Configuring Causes. *Pacific Philosophical Quarterly* 87: 141–162.
- Enoch, David. 2011. *Taking Morality Seriously: A Defense of Robust Realism*, 2011. Oxford: Oxford University Press.
- Harman, Gilbert. 1977. *The Nature of Morality*. New York: Oxford University Press.
- Hayward, Max Khan. 2019. Immoral Realism. *Philosophical Studies* 176: 897–914.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis* Oxford: Oxford University Press.
- Kim, Jaegwon. 1998. *Mind in a Physical World*. Cambridge: MIT Press.
- Kim, Jaegwon. 2005. *Physicalism or Something Near Enough*. Princeton: Princeton University Press.
- Lewis, David. 1988. What Experience Teaches. *Proceedings of the Russellian Society* 13: 29–57.
- ²² I would like to thank Robert Cowan, Roger Crisp, Gerry Dunne, Artur Harris, Alex Meyer and an anonymous reviewer for helpful comments.

- Luco, A. 2019. How Moral Facts Cause Moral Progress. *Journal of the American Philosophical Association*, 429–448.
- Majors, Brad. 2003. Moral Explanation and the Special Sciences. *Philosophical Studies* 113: 121–152.
- McLaughlin, Brian. 1992. The Rise and Fall of British Emergentism. In *Emergence or Reduction? Prospects for Nonreductive Physicalism*, ed. A. Beckermann, et al., 49–93. Berlin: De Gruyter.
- McPherson, Tristram. 2012. Ethical Non-naturalism and the Metaphysics of Supervenience. In *Oxford Studies in Metaethics*, vol. 7, ed. Russ Shafer-Landau, 205–234. Oxford: Oxford University Press.
- Papineau, David. 2001. The Rise of Physicalism. In *Physicalism and its Discontents*, ed. Carl Gillett and Barry M. Loewer, 3–36. Cambridge: Cambridge University Press.
- Parfit, Derek. 2011. *On What Matters*. Oxford: Oxford University Press.
- Railton, Peter. 1985. Moral Realism. *The Philosophical Review* 95: 163–207.
- Rosen, Gideon. 2020. What is Normative Necessity? In *Metaphysics, Meaning, and Modality: Themes from Kit Fine*, ed. M. Dumitru, 205–233. Oxford: Oxford University Press.
- Schiffer, Stephen. 1987. *Remnants of Meaning*. Cambridge, MA: MIT Press.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defense*. New York: Oxford University Press.
- Stringer, Ryan. 2019. Realist Ethical Naturalism. *Philosophical Studies* 175 (2): 339–362.
- Stringer, Ryan. 2021. Ethical Emergentism and Moral Causation. *Journal of Moral Philosophy* 18: 331–362.
- Sturgeon, Nicholas. 1985. Moral Explanations. In *Morality, Reason, and Truth*, ed. David Copp and David Zimmerman, 49–78. Totowa, NJ: Rowman and Allanheld.
- Sturgeon, Nicholas. 2003. Moore on Ethical Naturalism. *Ethics* 113: 528–556.
- Sturgeon, Nicholas. 2005. Ethical Naturalism. In *Oxford Handbook of Ethical Theory*, ed. David Copp, 91–122. Oxford: Oxford University Press.
- Wilson, J. 1999. How superduper does a physicalist supervenience need to be? *Philosophical Quarterly* 49: 33–52.
- Wilson, J. 2011. Non-reductive realization and the powers-based subset strategy. *The Monist* 94: 121–154.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.