



Epiphenomenalism for Functionalists

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Beebe, H. (2017). Epiphenomenalism for Functionalists. In B. Helen, C. Hitchcock, & H. Price (Eds.), *Making a Difference: Essays on the Philosophy of Causation* (pp. 286-306). Oxford University Press.

Published in:

Making a Difference

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Epiphenomenalism for Functionalists¹

Helen Beebee

Preprint. Please do not cite this version. The published version is:

‘Epiphenomenalism for Functionalists’ in H. Beebee, C. Hitchcock & H. Price (eds), *Making a Difference* (Oxford: OUP, 2017), 286-306

1. Introduction

The debate about the alleged incompatibility between non-reductive physicalism and the causal efficacy of the mental—often referred to as the ‘exclusion problem’ or ‘the problem of mental causation’—continues to rage. In recent years, that debate has focussed almost exclusively on the alleged problem posed by the non-reductive physicalist’s claim that mental properties are multiply realizable: in any given population or kind of being (humans, dogs, Martians, or whatever), or indeed perhaps in any given individual at different times, the property of, say, *being in pain*, call it *M*, might, for all we know, be realized by any of various neurological (or other physical) properties P_1, P_2, P_3, \dots . The problem is then supposed to be that ‘all the causal work’ is being done by the realizer property. Given the causal closure of the physical, it would seem that whichever of P_1, P_2 , etc. is instantiated is causally sufficient in the circumstances for the occurrence of a given effect *e* (saying ‘ouch’, say). Hence there is nothing left over for *M* to do: being in pain does not cause people to say ‘ouch’. And, since epiphenomenalism about mental properties is unacceptable, non-reductive physicalism must be abandoned. Non-reductive physicalists disagree with this conclusion, of course, and various robust defences of it have recently been offered.² I’m pretty confident that some such defence of the claim that multiple realization does

¹ Many thanks to Huw Price, Frank Jackson, Matt Tugby and Philip Pettit, as well as various seminar audiences, for many helpful comments. Special thanks are due—though they cannot now be delivered—to Peter Menzies, whose warm enthusiasm and support have been a constant throughout my philosophical career.

² See, for example, List & Menzies 2009, Raatikainen 2010, Weslake (f/c).

not preclude causal relevance succeeds. (If I had to put money on which, I'd plump for Woodward's defence, which relies on conceiving causation as difference-making (Woodward 2008, [this volume]); see §4 below.)

The vast majority of the recent literature focuses exclusively on the alleged problem of multiple realization. In doing so, however, it ignores a second—and, I think, less tractable—problem of mental causation, which remains even once we accept one of the available solutions to the problem of multiple realization. I shall call it ‘the Causal Role Problem’—although it is a very close relative of what has become known as ‘the Problem of Metaphysically Necessitated Effects’. The non-reductive physicalist takes mental properties to be multiply realized because mental properties are assumed to be *functional* properties. To be in pain, for example, is to be in some physical state or other, such that being in that state typically has certain causes (e.g. bodily damage) and effects (avoidance behaviour, utterance of expletives, etc.). Functional properties are defined in terms of their causes and effects. And that's where the problem lies: given a standard—and plausible—account of the individuation of events—one that can be reformulated as a plausible account of the kinds of properties that are apt for causal relevance—such properties are simply inapt for playing a causal role. *Qua* functional property, being in pain can no more be a cause of saying ‘ouch’ than (to use a well-worn example) having dormitive virtue can be a cause of sleep.³

Now, there are of course moves that can be made, which one might take to solve the Causal Role Problem. I argue in §5 that one such move (Antony 2008) fails, after explaining in a bit more detail what the Causal Role Problem is in §2, distinguishing it from the Problem of Metaphysically Necessary Effects in §3, and arguing that existing promising solutions to the Exclusion Problem are not solutions to the Causal Role Problem in §4. That leaves the non-reductive physicalist facing the allegedly unwelcome prospect of embracing epiphenomenalism about mental properties. As I argue in §6, however, it is far from clear that what Gabriel Segal (2009) calls

³ Authors who do discuss versions of this problem include Block (1990), Lyons (2006), Rupert (2006) and Segal (2009). What I say in this paper overlaps to some extent with what they say—although, as we'll see in §3, the focus is usually on the Problem of Metaphysically Necessary Effects rather than the Causal Role Problem.

‘epiphobia’ is warranted. Very many – perhaps even most – of our ordinary beliefs about the mental are entirely consistent with epiphenomenalism; in particular, epiphenomenalism is consistent with both the explanatory non-redundancy and the practical usefulness of mental properties, and with the causal efficacy of mental events.

2. The Causal Role Problem

We need, first, to distinguish between *events* on the one hand and *properties* on the other. Much of the mental causation literature proceeds as though properties are, or can be, the relata of token causation: we are generally explicitly asked to consider whether or not some mental property *M* is a cause of some physical property *P*, whereas in fact we are really being asked to consider a *particular* case, for example whether or not *my* being in pain, just now, was a cause of some subsequent bit of physical behaviour. Properties, just by themselves, make good candidates for the relata of a *general* causal relation (‘being in pain generally causes avoidance behaviour’, say), but not for the relata of a token one. At the token level, properties are most naturally seen not as the relata of causation but as candidates for causal *relevance*.⁴

Imagine, for example, that John says ‘hello’ to Jane as he walks into the office—as he often does—but on this occasion he says it in an unusually jaunty manner. Jane has two responses: she replies (‘hello, John’) in her customary fashion, but is simultaneously surprised at the jauntiness of John’s greeting. Intuitively, the jauntiness of John’s greeting is causally relevant to Jane’s surprise but not to her reply; after all, if John had greeted her in his usual, rather more dour, tone of voice, she would not have been surprised, but she would nonetheless have given the same reply as she actually gave.

⁴ I’m using ‘causal relevance’ in a pretheoretical way here. I briefly discuss Jackson and Pettit’s account of causal relevance (as opposed to what they call ‘causal efficacy’) in §6 below.

Counterfactual dependence, then, would seem to be a marker (at the very least) of the causal relevance of properties in cases of token causation. But what of the relation of token causation itself? For the purposes of this chapter, I shall assume the standard Lewisian story (Lewis 1986a). For Lewis, an event is a region of spacetime that has both essential and accidental properties, so that many events can (and normally do) occur in the very same spatio-temporal region, differing only in which properties are essential and which are accidental. Thus we can distinguish between two events, c_1 and c_2 , where *being a saying of 'hello' by John* and *being a jaunty saying of 'hello' by John* are the essential properties of c_1 and c_2 respectively. These are different (though not fully distinct) events, even though they occur in the very same spatio-temporal region, because there are possible worlds where c_1 occurs but c_2 does not—these being worlds where John says 'hello', but not jauntily. Similarly, Jane's reply (e_1) and her feeling of surprise (e_2) are distinct effects of John's behaviour, but again they are distinguished not by their spatio-temporal location (since they occur in the same spatio-temporal region, *viz.*, the region occupied by Jane at the time in question) but by their essential properties: e_1 is essentially Jane's utterance of 'hello, John', and e_2 is essentially her being surprised. So—given a counterfactual analysis of causation—it turns out that c_2 causes e_2 but not e_1 (had John not said 'hello' *jauntily*, Jane would still have replied but would not have been surprised). c_1 , by contrast (given some additional assumptions about the situation, at any rate), causes e_1 but not e_2 .

I shall assume that counterfactual dependence is sufficient for the causal relevance of properties in cases of token-level causation, and that the causal relevance of properties can itself be captured by appealing to the distinction between accidental and essential features of events. In effect, then, I am assuming that the instantiation of property F is (on a given occasion) causally relevant to the instantiation of property G if and only if there are events c and e , such that F is an essential property of c and G is an essential property of e , and c caused e . Thus the jauntiness (F) of John's saying 'hello' is causally relevant to the Jane's being surprised (G) because F is an essential property of c_2 , G is an essential property of e_2 , and c_2 caused e_2 . Intuitively, F is causally relevant to G because it makes a difference to whether or not some G -event occurs, and its making a difference consists in the fact that an essentially- F event causes an essentially- G event. By contrast, F is not causally relevant to whether or not Jane says

‘hello, John’, and this lack of causal relevance consists in the fact that c_2 is not a cause of e_1 .

Various aspects of the above story are, of course, open to dispute. The connection between the causal relevance of a property and the causal relation between events (or between whatever one thinks the relata of causation are) will vary depending on the account of causation and its relata that one adopts. Broadly speaking, since a difference-making approach to causal relevance (here conceived in terms of counterfactual dependence) seems easily the best bet, difference-making accounts of causation itself are going to be pretty closely aligned with some story about the causal relevance of properties. Things are going to be less straightforward for non-difference-making accounts of causation; but since our topic here is the causal relevance of properties, such accounts can safely be ignored. The real underlying point of assuming a Lewis-style account of causation is that it serves as a hook on which to hang the Causal Role Problem. But the problem will remain, so far as I can tell, whatever account of causation we sign up to.

The problem, then, is this: as Lewis notes (and given an abundant view of properties), not just *any* property is apt for featuring as an essential property of an event. If we are too permissive, we will find spurious counterfactual dependence, and hence spurious causation, between events; and functional properties—properties that are individuated according to causal roles—fall the wrong side of the line.

To illustrate the general idea, let’s consider three kinds of property that fairly obviously (to me anyway) give rise to spurious counterfactual dependence: disjunctive properties, dispositional properties, and what I’ll call ‘causally loaded’ properties. Let’s start with disjunctive properties. Suppose Jagbir smiles at Jake, and this makes Jake smile back. Jagbir’s smiling is also a smiling-or-an-ascent-of-Everest. But if we allow that an event occurs that has that disjunctive property as its essential property, then we’ll get spurious counterfactual dependence: assuming that possible worlds where Jagbir climbs Everest are very distant from the actual world, the closest world where Jagbir fails to smile-or-climb-Everest is just the closest world where she

fails to smile. So Jake's smiling counterfactually depends on that event, as well as on the more mundane event whose essential property is Jagbir's smiling. But that is surely spurious dependence: Jagbir's smiling causes Jake to smile back, but her smiling-or-climbing-Everest does not.⁵

Dispositional properties similarly give rise to spurious dependence. The classic example is, of course, Molière's *virtus dormitiva*. If we count ingesting a soporific (that is, something with the disposition to induce sleep), as well as ingesting something with the categorical basis of that disposition, as a cause of someone's falling asleep on the grounds that their falling asleep counterfactually depends on their having ingested a soporific, we are clearly double-counting: the counterfactual dependence of falling asleep on having ingested a soporific is spurious. If that doesn't sound obvious, consider the fact that pretty much any time we have a true causal claim of the form '*c* caused *e*', the laws of nature together with some additional facts about the circumstances surrounding the occurrence of *c* will, along with the occurrence of *c* itself, entail (or perhaps merely make likely) the occurrence of *e*. This being so, dispositions are extremely cheap. Pillowcases, for example, have the disposition to move in such-and-such a manner (the way mine are in fact currently moving) when hung on the line just so and exposed to exactly the strength and direction of wind that they're currently being exposed to; pigeons relevantly similar to the one I'm now observing have the disposition to flap about in circumstances that are exactly similar to those that are currently causing the pigeon I'm observing to flap about; and so on. To count such dispositional properties as causally relevant to the movement of my pillowcases or the flapping of the pigeon would seem to be double-counting if anything is.⁶

⁵ Lewis doesn't ban disjunctive properties all together; just those that are 'overly' disjunctive; see his 1986a, §VIII. One might try to argue that mental properties evade the Causal Role Problem by virtue of being (a) disjunctions of physical properties, but (b) not 'overly' disjunctive. That's a possibility I consider in §5.

⁶ Some philosophers—in particular, dispositional essentialists—claim that not all dispositions have categorical bases (e.g. Molnar 2003, Bird 2007), or, more strongly, that since there are no genuinely categorical properties, *all* dispositions lack categorical bases. Such views escape the argument just presented. Nonetheless, it's not obvious that all dispositionalist views escape an argument in the same general ballpark. In particular, *prima facie* at least, non-fundamental dispositions, such as fragility or being a hallucinogen, would still seem to

Third, consider properties like *being a cause of*, *being a potential cause of*, and *being a likely cause of*—call these ‘causally loaded’ properties. The seminar currently going on in the next room, *e*, is caused by many and various other events, each of which instantiated one or more properties that were causally relevant to the seminar’s occurrence. Each of them, trivially, instantiated the properties *being a cause of the seminar* and *being a potential cause of the seminar*; and some (but perhaps not all) of them also instantiated the property *being a likely cause of the seminar*. Are those properties of the various causes of the seminar *themselves* causally relevant to the seminar? Surely not. Again, counting causally loaded properties as themselves bearers of causal relevance is surely double-counting if anything is.

While there are differences between functional properties on the one hand and dispositional and causally-loaded properties as just conceived on the other, there is enough commonality between them to make it clear that to count functional properties as genuine bearers of causal relevance would, as in these other cases, be double-counting. Consider *being in pain* again, where to be in pain is to be in some physical state or other, such that being in that state typically has certain causes and effects. In relevant respects at least, such a property is indistinguishable from many run-of-the-mill dispositional properties. *Being a hallucinogen*, for example, is multiply realized (there are many substances that typically cause hallucination), and is explicitly defined in terms of its (typical) effect.

Being in pain is of course defined in terms of *both* its typical causes and its typical effects. But it’s hard to see how that’s going to help. Let’s hoke up a new case in order to drive the point home. Define what it is to be a *letter-maker* as to be in a state that typically causes a letter of the alphabet to appear on my computer screen, *and* is typically caused (at least in part) by my presence at the computer. I have been instantiating that property, off and on, for the last hour or so. So, currently, is Fang,

generate the double-counting worry—though a position like Shoemaker’s subset view (2001) might evade this worry.

who is wandering on my desk looking for some affection and, as it happens, standing on the ‘k’ key with his paw—something that typically (and indeed on this occasion), but not invariably, causes a ‘k’ to appear on the screen. *Being a letter-maker* is, at present, a property of Fang, and it generates counterfactual dependence between the event essentially specifiable as Fang’s currently being a letter-maker (*c*)—if there is such an event—and a ‘k’ appearing on my screen (*e*): in current circumstances, had *c* not occurred, nor would *e*. But *being a letter-maker* is clearly not a property that is genuinely causally relevant to the ‘k’ appearing on the screen.

If we were to grant causal relevance to functional properties, then, we would have no reason not to grant causal relevance to a whole host of other properties—properties that (so I claim) *manifestly* lack causal relevance. The lack of causal relevance of dispositions, in particular, is (excepting views according to which dispositions can be fundamental) pretty widely accepted. (See for example Prior, Pargetter and Jackson 1982; Lewis 1986a, 268; Pettit, THIS VOLUME, PAGEREF.) The fact that we can, in addition, hoke up disposition-like properties for any given instance of token causation (as in the pillowcase and pigeon examples) adds grist to that particular mill. The argument for the Causal Role Problem, then, is that we cannot countenance genuine causal relevance for functional properties (or, equivalently, we cannot countenance such properties as essential features of events) without allowing unacceptable double-counting.⁷

I conclude that, given some pretty standard and apparently plausible assumptions about causal relevance, the standard version of non-reductive physicalism – *viz.*, functionalism – entails epiphenomenalism about mental properties.

⁷ One might wonder at this point whether the Causal Role Problem is just the Exclusion Problem under another name. It isn’t; see §4 below.

3. The Causal Role Problem and the Problem of Metaphysically Necessitated Effects

As I said earlier, the Causal Role Problem is not a new problem; in fact, it is a very close relative of what Richard Rupert (2006) calls ‘the Problem of Metaphysically Necessitated Effects’. The problem articulated above is that functionally specified properties cannot do duty as essential properties of events (equivalently: cannot be bearers of causal relevance)—since, if they did, they would generate non-causal counterfactual dependence relations. The Problem of Metaphysically Necessitated Effects (hereafter PMNE) is that, since whether a physical property *P* counts as a realizer of *pain* depends upon its being a property that has certain effects (some event of kind *G*, say), the instantiation of *G* is metaphysically necessitated by the instantiation of *pain*. This violates the Humean claim that causal relations are contingent.⁸ As Rupert puts it:

Functionalist mental properties are individuated partly by their relation to the very effects those properties’ instantiations are thought to cause.

Consequently, functionalist causal generalizations would seem to have the following problematical structure: The state of being, among other things, a cause of *e* (under such-and-such conditions) causes *e* (under those conditions).

The connection asserted lacks the contingency one would expect of a causal generalization. (2006, 256)

PMNE and the Causal Role Problem clearly have the same general shape. They are not, however, quite the *same* problem – and I think the latter is a harder problem than the former. Note that PMNE gains its force from the idea that a functional property is (as Rupert puts it) ‘the state of being ... a cause of *e*’—and hence the having of that property cannot itself cause (be causally relevant to) *e*. Other authors who have raised the general problem posed by the lack of distinctness between functional properties and their alleged effects have, in effect, also been raising a version of PMNE: the

⁸ The Humean claim of course can be, and has been, denied. See Rupert 2006, 258-60, for arguments that denying it is not a promising way to go in the context of the problem under discussion.

problem, they claim, is the entailment relation that holds between the two (see Ludwig 1994 and Lyons 2006).⁹ But entailment is a stronger relation than what is needed to generate spurious counterfactual dependence, which is what the Causal Role Problem is concerned with. Recall an example from §2: Jake's smiling counterfactually depends on Jagbir's smiling-or-ascending-Everest—but the former does not *entail* the latter. Instead, what generates the dependence is (a) the entirely contingent *counterfactual* dependence of Jake's smiling on Jagbir's smiling, and (b) the fact—again, an entirely contingent fact—that worlds where Jagbir climbs Everest are much further away from actuality than are ones where she fails to smile, and so the closest possible world where Jagbir does neither is simply the closest possible world where she fails to smile.

Moreover, the standard functionalist specification of mental properties, like the property of being a hallucinogen, is considerably looser than that assumed by proponents of PMNE. Take the case of pain again. First, whether we take the definition of the 'pain role' to be a matter for conceptual analysis (analytic functionalism) or for fleshing out by appeal to our best scientific, psychological theory (psycho-functionalism), or whatever, no remotely plausible specification of pain's definitive causes and effects is going to be precise enough to establish a metaphysically necessary connection between being in pain and exhibiting any maximally specific kind of behaviour. The precise kind of behaviour elicited by being in pain will vary enormously between species, between individuals, and even between different occasions for the same individual. Come at me with a needle while we're in the pub and I'll rapidly remove myself from your vicinity; my behaviour when having a blood test is quite different. Having pain inflicted on one might elicit a string of expletives, or a simple 'ouch, that hurts!', or merely moaning or yelping (or indeed, in the doctor's surgery, no more than a slight wince). The most that any plausible specification of the definitive effects of pain will necessitate is that one exhibit some form of behaviour of a very general kind; it will not necessitate any maximally specific form of behaviour that a given person (or animal) manifests on a given occasion.

⁹ A notable exception is Lewis 1986a, as we'll see in the next section.

Second—and more problematically for PMNE—any plausible definitive specification of the effects of pain is going to be hedged: there’s going to be a ‘typically’ in there, or perhaps a ‘ceteris paribus’. On a good day I can manage a blood test without even so much as a mild wince. (It still hurts, though. I’m just pretending that it doesn’t.) As we saw with the case of *being a hallucinogen*, it’s entirely conceptually and metaphysically possible that a particular person, on a particular occasion, is in pain and nonetheless fails to exhibit the typical behaviour that partially defines what it is to be in pain.

It is therefore unclear whether PMNE itself really *is* a problem for functionalism. First, even if we assume that *some* form of, say, avoidance behaviour is necessitated by being in pain, the fact that the specific behaviour exhibited is not metaphysically necessitated may be enough to avoid the problem. Second, and more seriously, once we note the presence of the ‘typically’ in our functional specification, again we lose metaphysical necessitation: if avoidance behaviour is only *typically* caused by being in pain, then such behaviour is not metaphysically necessitated by my being in pain.

The Causal Role Problem remains, however, since nothing in the argument of §2 depended on any necessitation between the instantiation of the property in question and the occurrence of the effect. Recall Fang, who, while wandering on my desk, acquired the property of *being a letter-maker*. His instantiating this property did not metaphysically necessitate the appearance of a ‘k’ on my screen (*e*): in the circumstances, *e* might not have occurred, consistent with Fang’s being a letter-maker in the right circumstances. For he might have trodden on a different key, or he might have trodden on the ‘k’ key but failed thereby to produce a ‘k’ on the screen.

4. Who is the Causal Role Problem a problem for?

4.1 Role functionalism vs. realizer functionalism

A distinction is often made between ‘role’ functionalism on the one hand and ‘filler’ or ‘realizer’ functionalism on the other. According to role functionalism, a mental

term such as ‘pain’ rigidly designates a second-order property—the property of having such-and-such causal role—whereas according to realizer functionalism, a mental terms nonrigidly denotes the first-order, physical property that realizes that causal role (see e.g. McLaughlin 2007, 151-2; Bennett 2007, 323). Realizer functionalism is incompatible with multiple realization: if more than one physical property actually plays the pain-role, then pain cannot be *the* physical property that plays the pain-role. However, following Lewis (1980), we might relativize the concept of pain to different species in order to account for the possibility of, say, a Martian, for whom (thanks to a very different physical make-up to us) some entirely different state occupies the pain role. Thus Lewis endorses the claim that ‘*X* is in pain *simpliciter* if and only if *X* is in the state that occupies the pain role for the *appropriate* population’ (1980: 219), where the appropriate population would be, say, normal human beings in my case, and normal Martians in the case of our Martian.

It is often said that, since realizer functionalism is, in effect, a version of the type-type identity theory (since on this view pain-for-humans, say, just *is* the firing of C-fibres, or whatever), it is immune to the Exclusion Problem (Kim 1989, Bennett 2007, McLaughlin 2007). Once we grant that it is pain-for-humans (or perhaps something even more relativized than this—see Kim 1989, 38)—call this pain*—that is our candidate for causal relevance, there is no problem of a competition for causal relevance between mental and physical properties, since a property cannot be in competition with itself. If this a convincing response to the Exclusion Problem, then of course it looks as though the same will be true of the Causal Role Problem. If, in saying that my being in pain* caused me to wince, I am merely referring to whatever property *P* realizes pain*, then—assuming that *P* was causally relevant to my wincing—my being in pain* really did cause it.

If this constitutes a solution to both problems, then so be it. Since the kind of functionalism that allegedly solves the problem is a version of the type-type identity theory, it is (as Kim (1989: 39) notes) a reductionist position—and hence it is no help to genuinely nonreductive physicalism, which is the position I’m interested in here. For what it’s worth, however, I’m not so sure realizer functionalism *does* solve the

Causal Role Problem. If mental terms only non-rigidly refer to the physical role-players then while, in the actual world, mental terms are not multiple *realized*, they are nonetheless multiply *realizable*. After all, human beings could have evolved differently, and some other physical property could have ended up playing the pain*-role—and then *that* property, and not *P*, would have been pain*. So the ‘identity’ between pain* and *P* is mere contingent identity. This being so, it seems coherent to ask whether being in pain* gets to be causally relevant (to my wincing, say) by virtue of being the physical property it is or by virtue of playing the causal role that it does. And it’s hard to see, given the argument of §2, how the answer could be ‘both’. Lewis himself concurs:

Whenever some term nonrigidly designates the occupant of a role, and that role could be occupied in a variety of ways, the term becomes unsuitable for essential specification of events. If being fragile means having some or another basis for a disposition to break when struck, and if many different properties could serve as such bases (under this or otherworldly laws), then no genuine event is essentially classifiable as the window's being fragile. There is a genuine event which is accidentally classifiable in terms of fragility; essentially, however, it is a possession of such-and-such molecular structure, that being the actual basis of the window's fragility ... And if I am right to think that mental states are definable as occupants of causal roles, then no genuine event is essentially classifiable as my being in pain. There are pain events, no doubt of it; but they are pain events only accidentally, just as pain itself is a property that only contingently occupies its role and deserves its name. Essentially, the events are firings of neurons, perhaps—unless ‘firing’ and ‘neuron’ also are terms for occupants of roles, in which case we must get more physical before we finally reach an essential classification. (Lewis 1986a: 268)

4.2 Counterfactual-based solutions to the Exclusion Problem

Here is a line of thought that has motivated many recent attempts to solve the Exclusion Problem (see e.g. List and Menzies 2009, 489). The idea that the causal

sufficiency of the physical precludes the mental from having any causal status is grounded in a ‘production’ conception of causation, of which paradigmatic cases would be things like shooting people and the collision of billiard balls: cases where there is a localized process or transfer of some entity or quantity (a bullet, energy-momentum, etc.). On such a conception of causation, it looks as though P (a physical property that is causally sufficient in the circumstances for some effect e) and M (a supervening mental property) could only *both* cause e if each, separately, was involved in some sort of productive process culminating in the effect, as with two assassins independently shooting the victim at the same time, or my reaching for the aspirin because I have simultaneously stubbed my toe and banged my head. But since widespread overdetermination is unpalatable (and, in any case, overdetermination of the kind just described is manifestly nothing like what happens in putative cases of mental causation), the causal inefficacy of the mental follows.

But we can reject this line of thought. A way of thinking about causation that is considerably more conducive to accommodating the mental is as a matter of *difference-making*, which we can define in terms of—or at least legitimately take to be very closely related to—counterfactual dependence. Once we make this basic move, the Exclusion Problem starts looking a lot more tractable. After all, it’s pretty uncontroversial that plenty of things counterfactually depend on the instantiation of mental properties: if I hadn’t been in pain I wouldn’t have taken the aspirin, if I hadn’t wanted a beer I wouldn’t have ordered one, if I hadn’t believed that today was Wednesday I wouldn’t have put the rubbish out, and so on.

I myself am inclined to think that responses to the Exclusion Problem along the lines that Woodward, List and Menzies, and others have pursued are pretty promising—or rather, and here’s the rub, they *would* be if it weren’t for the fact that mental properties are conceived by these and similar solutions as functional properties. As I said right at the beginning, the feature of nonreductive physicalism for which the Exclusion Problem is (at least *prima facie*) a problem is multiple realization: it is the problem that multiply realizable properties (such as functional properties) would seem to have no additional causal ‘work’ left for them to do, since quite enough work is

being done by the realizer properties on their own. The Causal Role Problem is a more basic problem, in that the reason *why* nonreductive physicalists take mental properties to be multiply realizable in the first place is that they are, precisely, functional properties. The problem is that, without some restrictions in place concerning what properties are apt for counting as essential properties of events—or, equivalently (I have assumed), concerning what properties can count as causally relevant—counterfactual dependence of the kind identified by difference-making solutions to the Exclusion Problem, just by itself, is too permissive as a criterion for causal relevance and hence genuinely causal difference-making. And some pretty intuitive considerations indicate that whatever the right restrictions are, functional properties are going to get banned.

5. Role properties vs. disjunctive properties

Here is a potential objection to my insistence that the Causal Role Problem is a genuine problem that remains even given a counterfactual-style solution to the Exclusion Problem. The first part of the objection runs like this. Sure, our mental *concepts* are defined in terms of their causes and effects; our *concept* of pain is the concept of a state that typically has such-and-such causes and effects. But the property that the concept picks out is really a disjunction of the various realizers of that causal role—and that property *itself* is not, as it were, inherently causally-loaded. After all, it's a substantive empirical claim that a given disjunction of physical properties typically has such-and-such causes and effects. With this move on the table, the Causal Role Problem, if it is still a problem, applies not because mental properties *themselves* are functional, but because they are disjunctive. And the second part of the objection runs as follows. While, as we've seen, Lewis himself bans properties that are 'too disjunctive' from serving as essential properties of events because they give rise to spurious counterfactual dependence (as in the example of Jagbir smiling-or-climbing-Everest), there is room for a more nuanced approach: there is a way of ruling out hoked-up disjunctions while making room for the kind we want to allow. Causal Role Problem solved.

This objection depends on two claims: the one about the disjunctive (as opposed to functional) nature of mental properties, and the one about distinguishing between hoked-up disjunctions and the ‘nice’ kind, such as the disjunctions that (allegedly) constitute mental properties. I’ll briefly discuss the first claim before arguing that the second is false.

The obvious *prima facie* problem with the claim that mental properties can be identified with disjunctions of physical properties is that in effect gives up on nonreductive physicalism—the position at issue in this chapter—by identifying mental and (albeit wildly disjunctive) physical properties, which of course is what reductive, type-physicalists do (see §3 above). As Louise Antony (2008) points out, however, we should be wary of the claim that identifying mental properties with disjunctions of physical properties really amounts to an endorsement of reductive physicalism. As she notes, on a conception of properties according to which they are simply sets of possible worlds (so that the property *F* just *is* the set of worlds such that at least one proposition ascribing *F* to something is true at each member of the set), a consequence of the view under discussion is that the mental predicate ‘*M*’ and the corresponding disjunction ‘*P*’ do indeed refer to the very same property: $M = P$. However, she claims that this does not make the view she is defending turn out to be simply a version of the identity theory. Granted a possible-worlds conception of properties, there aren’t really such things as disjunctive *properties*—a set of worlds is just a set of worlds, after all, and there’s nothing inherently disjunctive about a set of worlds—so there are really only disjunctive *predicates*. And, given physicalism, *any* mental predicate is going to pick out some set of worlds describable in purely physical terms. So, on pain of vacuity, we should not think of the truth of $M = P$, just by itself, as a lapse into the type identity theory—that is, into reductive physicalism. Rather, we should think of what Antony calls ‘strong reductionism’ as the thesis that ‘every mentalistic predicate is necessarily co-extensive with some *proprietary* predicate of lower-order or lower-level science’ (2008: 173). Since the kind of unwieldy disjunctive predicate that is (on Antony’s view) co-extensive with a given mentalistic predicate will not *itself* be a proprietary predicate of physical science (even if each disjunct is such a predicate), Antony’s view denies strong reductionism, and hence counts as a version of nonreductive physicalism.

Let's assume, then, that nonreductive physicalism is indeed compatible with taking mental predicates to pick out the same properties as disjunctive physical predicates. Now, according to Antony every higher-order, mentalistic predicate 'is necessarily co-extensive with some lower-order, possibly infinitely long, disjunctive predicate' (2008: 170). But the 'necessarily' part surely cannot be right. For example, there may well be other possible worlds where human beings evolved in such a way that some proprietary physical predicate ' P_1 ' refers to a physical property that realizes pain in that world, but that that physical property is *actually* a realizer of the *mildly pleasurable tickling sensation* role. In that case, ' P_1 ' cannot be among the disjuncts of our unwieldy disjunctive predicate ' P ', since if it were, it would follow that an *actual* human being with property P_1 is in pain rather than undergoing a mildly pleasurable tickling sensation. So, if the view under consideration here is to have any prospects, it looks as though we need to deny what Antony asserts and hold that any mental predicate is *in fact* co-extensive with some lower-order disjunctive predicate. Or, to put it another way, we need to hold that our mental predicate ' M ' nonrigidly refers to property P , where ' P ' is a disjunction of proprietary predicates of some lower-order science.

Of course, if the suspicion raised earlier (and endorsed by Lewis), that mental terms that are nonrigid designators of physical properties cannot serve as essential specifications of events, is right, then the same point applies here too. But let's leave that aside and move on the second claim that needs to be established, *viz.*, that there is a way of allowing the 'nice' disjunctive properties picked out by mental terms to count as causally relevant while banning hoked-up properties such as smiling-or-ascending-Everest.¹⁰

The first thing to note is that nothing in recent, difference-making attempts to solve the Exclusion Problem helps us to discriminate between hoked-up and nice disjunctions. As we saw in §2, the fact that hoked-up disjunctions can deliver

¹⁰ As we've just seen, Antony cautions against thinking in terms of disjunctive *properties*. I'm doing so here for ease of exposition; nothing hangs on it.

counterfactual dependence, and so ‘make a difference’ in that sense, is part of the problem and not part of the solution. Similarly, nothing in the interventionist solution to the Exclusion Problem will do the trick. In almost all circumstances, intervening on whether or not someone is smiling-or-ascending-Everest generates exactly the same pattern of dependence as does intervening on their smiling alone, since—for almost everyone all the time, and for a very few intrepid people almost all the time—the intervention will proceed by inducing the person in question to smile or not, ascending Everest not being a viable option. So hoked-up disjunctive properties can happily satisfy Woodward’s conditions for causal relevance.

This, of course, is not an objection to List and Menzies’ or Woodward’s accounts *qua* solutions to the Exclusion Problem; it merely makes the familiar point that independent constraints on what can count as a causally relevant factor or an admissible value of a variable are needed in order to rule out spurious cases of dependence. For the most part, we can assume standard restrictions along something like the lines described in §2 above. But of course the point of the Causal Role Problem is precisely that such restrictions throw out mental causation along with the bathwater. Is there such a constraint to be had? Antony argues, in effect, that there is.

The question, as Antony puts it, is, ‘what makes it the case that some disjunctive predicates express nomic properties, while others do not?’ (2008: 169). ‘A property is nomic if it participates in lawful objective regularities’, Antony says (2008: 170), which is perhaps not as precise as we might like for current purposes, but nomicity is certainly in the same ballpark as causal relevance; in any case, the basic contrast Antony is interested in is with hoked-up disjunctive properties (see 2008: 167), which is what I’m interested in here. So let’s take the notion of a ‘nomic property’ to be sufficiently well understood, at least for now.

Antony’s answer to the question just posed appeals to Nelson Goodman’s notions of entrenchment and projectibility. First, some definitions:

(1) ‘Entrenchment’ is an observable socio-linguistic property ... (2) A *predicate* will be said to be ‘projectible’ just in case it (a) is entrenched in some community and (b) can in fact be used to state correct predictions and robust (although possibly *ceteris paribus*) generalizations. (3) A *property* will be said to be projectible if and only if it is expressed by some projectible predicate, in some language, for some intentional beings. (2008: 170)

With this on the table, Antony says:

typically, but not necessarily, entrenched predicates will be projectible. That is, predicates that are entrenched permit and will continue to permit the formulation of correct predictions and robust generalizations ... The explanation for the projectibility of a predicate, and hence, in many cases, for its entrenchment, is that the property expressed by that predicate is nomic. Finally, all projectible properties are nomic, but not all nomic properties need be projectible. There may be nomic properties that neither we, nor the members of any other linguistic community, are ever able to express by means of a projectible predicate. (2008: 170)

The basic idea, then, is that mentalistic predicates are (normally) entrenched precisely *because* they are projectible: they ‘permit the formulation of correct predictions and robust generalizations’. And the reason for the projectibility of mentalistic predicates, in turn, is that the properties they express are nomic. Thus, for example—since we know that some mentalistic predicate ‘*M*’ (‘pain’, say) is both entrenched and projectible—we can safely infer that the property it expresses is nomic. The corresponding wildly disjunctive physical predicate ‘*P*’, while co-extensive with ‘*M*’, is neither entrenched nor projectible: it fails to permit the formulation of correct predictions and robust generalizations. Nonetheless, the *property* that ‘*P*’ expresses—being the very same property as that expressed by ‘*M*’—*is* nomic.

Antony's implied answer to the question about the distinction between predicates that do and don't express nomic properties, then, is that the non-nomic properties fail to 'participate in lawful objective regularities'. But now we face a problem: if properties picked out by functionalist, mentalistic predicates fall on the nomic side of the nomic/non-nomic divide, why should we not say the same for dispositional and causally-loaded predicates, and for at least some disjunctive ones? After all, such predicates can in principle be—and many of them are—both entrenched and projectible. Consider predicates like 'hallucinogen' and 'fatal'. These predicates feature in perfectly good robust and stable (if sometimes *ceteris paribus*) generalizations: people who suffer fatal accidents die, and people who take hallucinogens generally end up hallucinating. So Antony's view faces a dilemma: if *being a hallucinogen* and *being fatal* are nomic properties, then the causal relevance of the mental is secured at the price of pervasive double-counting. On the other hand, if they are not nomic properties, then the entrenchment and projectibility of a predicate fails to license the inference to the claim that the property the predicate expresses is nomic.

My view, unsurprisingly, is that we should accept the second horn of the dilemma. Hallucination and death are things that we are generally interested in—and so it's entirely sensible to have a general term covering the properties that are liable to cause them, however multifarious those causes might be. (If we weren't interested in hallucinogenic or fatal properties of things, we wouldn't have invented the words.) But *being fatal* is not causally relevant to death: it is a conceptual truth, and not an empirical discovery, that fatal accidents cause death. Similarly for the fact that hallucinogens often cause hallucination: they wouldn't be hallucinogens if they didn't. And similarly for mentalistic predicates.

Of course, the view under discussion here distinguishes between mentalistic *predicates* and the properties (expressible in physical terms only by using wildly disjunctive predicates) that they refer to. One might try to argue that it is therefore entirely appropriate to think of the properties that are expressed by the kinds of predicate I'm interested in as genuinely nomic. I think such an argument fails.

Suppose we think of *being fatal* as a property that is expressible in terms that don't refer to its effects only by using a wildly disjunctive and open-ended predicate: 'being a head-on high-speed collision, being an airliner crash, being the grabbing of a live electricity cable, being the ingestion of large quantities of cyanide, ...', say (call this property *P*). *P* is only fully expressible by actual, finite and non-omniscient human beings by means of the predicate 'fatal'. Granted, it is *not* a conceptual truth that *P* causes death (there are, presumably, distant worlds where drinking cyanide is good for you, grabbing live electricity cables delivers a pleasant tickling sensation, and so on). But that doesn't make it any more plausible to claim that *P* is a nomic property. That there are (extremely!) stable generalizations that involve a predicate that expresses *P* simply doesn't provide us with any grounds for making that claim, since the stability of those generalizations is explained entirely by the fact that the predicate that determines the extension of *P* ('fatal') is a dispositional predicate. And, again, the same point applies to mentalistic predicates if they are conceived in functionalist terms.

I conclude that conceiving mental, functional predicates as co-extensive with physical, disjunctive predicates that refer to physical properties does not deliver a way of granting mental properties causal relevance.

6. Is epiphenomenalism really so bad?

The point of this chapter so far has been to justify taking the Causal Role Problem seriously. Doubtless the argument is not decisive, and perhaps the problem can be solved. But suppose it can't be solved. The result—given our starting assumption, *viz.*, the truth of functionalism—would be epiphenomenalism.

It's generally assumed that epiphenomenalism with respect to the mental would be a completely unpalatable result.¹¹ But how bad *would* epiphenomenalism be, really?

¹¹ Three exceptions: first, Jack Lyons defends epiphenomenalism in his 2006, §§3 and 4. The point I make below about the causal efficacy of mental events is basically a short version of Lyons' claim that 'property epiphenomenalism' does not entail 'event epiphenomenalism'. Lyons argues that (property) epiphenomenalism is not only acceptable but a virtue of non-

Not nearly as bad as one might think, in fact—or so I shall argue. I'll argue that the kind of epiphenomenalism that the Causal Role Problem points to, while it does undermine the causal relevance of mental properties, leaves much of what we want to say about the mental intact: the causal irrelevance of the mental does not entail that mental properties are explanatorily redundant, nor does it entail that mental *events* are causally inefficacious. Finally, it does not entail that conceiving the world in terms of mental properties is useless for the purposes of controlling ourselves and others.

Let's start with the explanatory point. Here we can appeal to a Lewisian story about causal explanation (Lewis 1986b): to explain an event is to provide information about its causal history. Functional mental properties can perfectly well satisfy this requirement. When I explain your behaviour *B* (saying 'ouch!' and grimacing, say) by citing the fact that you are in pain, I *do* provide information about the causal history of *B*. In particular, I provide the information that *B* was caused by some event with some physical feature or other that typically plays the pain role—even though that role might well include, precisely, exhibiting *B*-like behaviour.

One might object that such an 'explanation' is trivial, since in effect it amounts to no more than explaining some event *e* by saying that was caused by some event or other of a kind that typically causes *e*—which hardly sounds like front-page news. Well, our explanation here may not be front-page news, but it is still informative, since *B* *could* have been caused by an event with some physical feature that does *not* typically play the pain role. You might, for example, have wanted to deceive me into thinking that you were in pain because you were looking for sympathy, and said 'ouch' and grimaced for that reason. The physical features upon which that desire supervenes do not typically play the pain role. Or you might have been acting in a play that called for pain behaviour at the moment in question, or responding to someone who has just threatened to kidnap your cat unless you say 'ouch' and grimace right now. Again, in

reductive physicalism. I lack the space to discuss this interesting suggestion here. Second, Segal (2009) also makes the point about the event/property distinction and adds his own defence of epiphenomenalism in response to a version of the Exclusion Problem. Finally, Frank Jackson (2012: §VII) argues that objections to epiphenomenalism from introspection, evolution and knowledge are unsound.

such cases the physical state in question is not one that typically plays the pain role. So citing the fact that you were in pain rules out all of these possible causal histories, just as it would if being in pain were genuinely causally relevant to *B*.

This is an apt place to compare the kind of epiphenomenalism under discussion here with Jackson and Pettit's distinction between 'programme' and 'process' explanation. According to Jackson and Pettit's original account, 'properties may be causally explanatory properties without being causally productive or efficacious ones. These properties programme the result to be explained, rather than actually bringing it about, and are the properties appealed to in what we called programme explanations' (1988: 400). And the kinds of property that Jackson and Pettit want to count as being 'causally explanatory' without being 'productive or efficacious' include functional properties.

Jackson and Pettit's 1988 view is, I think, compatible with epiphenomenalism about mental properties. Indeed, some of what they say about programme explanations seems to suggest that at least some properties that can crop up in programme explanations are genuinely epiphenomenal. Thus they say:

We may explain the conductor's annoyance at a concert by the fact that someone coughed. What will *actually have caused* the conductor's annoyance will be the coughing of some particular person, Fred, say; when we say that it was someone's coughing that explains why the conductor was annoyed, we are thinking of someone's coughing as Fred's coughing or Mary's coughing or Harry's coughing or ..., and saying that any of these disjuncts would have caused the conductor's annoyance—it did not have to be Fred. (1988: 394, my italics)

The implication here is that *someone's* coughing did *not* cause the conductor's annoyance, despite the legitimacy of the explanation described above. So it looks as

though we can have genuine (programme) explanation in the absence of causal relevance on Jackson and Pettit's 1988 view.

So far, so good. In later work, however, Jackson and Pettit (1990) distinguish not just between those properties that feature in programme and process *explanations*, but between 'causally efficacious' and 'causally relevant' properties—where their response to the Exclusion Problem is roughly to argue that the problem trades on the assumption that causal efficacy is the only kind of causal relevance there is, and that this assumption is false. It is this further move that I think needs to be rejected: to say that a property is causally *relevant* is, it seems to me, to ascribe a distinctively causal, and not *merely* explanatory, role to it. And, as I've argued, no such role can be ascribed to functional properties.

In his contribution to this volume, Pettit attributes a 'distinct-existences assumption' to List and Menzies' account of mental causation: 'higher-level [multiply-realized] properties—or more strictly, their instances—are distinct existences ... from the properties or property-instances to which they bear [significant law-like] relations' ([PAGEREF]). Pettit here clearly has in mind the Problem of Metaphysically Necessitated Effects—a problem which, as we saw in §3, connects with concerns about the failure of the principle that causes and effects are distinct. (Thus: the disposition to dissolve in water 'is not a distinct existence from the dissolving and not capable, for example, of bearing a causal relationship to it: that is, a contingent relationship that might not have obtained, even under presumptively suitable conditions' (this volume, [PAGEREF].)

Pettit further claims that the 'distinct-existences' assumption is, in fact, made in his and Jackson's earlier work with respect to functional properties. In a footnote, he says:

While we discussed the application of the model to cases of dispositions, where the distinct-existences assumption does not apply, we used that

application only to illustrate how the programming relationship need not add to our causal understanding. Notice that the distinct-existences assumption does apply, however, with functional as distinct from dispositional states. Unlike dispositions these are not characterized by just one manifesting connection but by the fulfilment of a number of conditions—usually, typical rather than invariable conditions. When the existence of the state requires a number of conditions to obtain, then even if they invariably include the connection with the effect to be explained, invoking the state in the explanation directs us to a connection between the other conditions required for the state to obtain and that effect. (Pettit, this volume, [PAGEREF] n.6)

Pettit's point here, then, is that, thanks to differences between dispositional and functional properties, the latter properties (or perhaps instances thereof), but not the former, are distinct from their effects. Thus, while the 'view that dispositions can be difference-makers and causes' is 'utterly implausible' (this volume, [PAGEREF, n.5])—and hence, I take it, the view that dispositions are bearers of causal relevance is similarly implausible—the same cannot be said of functional properties.

In fact, this point is similar to the point made in §3 that PMNE doesn't really apply to functional properties (or at least not the ones that nonreductive physicalists take to be mental states). But of course that all still leaves the Causal Role Problem in the running, since many of the properties that lead to double-counting are ones such that are 'distinct existences' from their putative effects in Pettit's sense: the putative causal relationship between them and their effects *is* 'a contingent relationship that might not have obtained, even under presumptively suitable conditions'. The sending of an invitation to the speaker, for example, was a *potential cause of the seminar* going on next door, *e* (call this property *F*), since it was an actual cause of *e*; but the relationship between the cause-event's being *F* and the occurrence of *e* is contingent: a last-minute train cancellation might have prevented the speaker from turning up, for example. Similarly, the relationship between disjunctive properties and events that counterfactually depend on them is not one where the distinct-existences assumption fails; the contingency of the relationship between Jagbir's smiling-or-ascending-

Everest and Jake's smiling is every bit as contingent as the relationship between Jagbir's smiling and Jake's smiling. So we are still left with the conclusion that functional properties cannot, *pace* Pettit, be bearers of causal relevance.

That said, it seems to me that the general spirit of Jackson and Pettit's account is, in places, closer to the epiphenomenalist position currently under discussion than Pettit now wants to accept. For one thing, several of the cases of programme explanation that they mention seem to me to count as genuinely explanatory, for the reasons they give, independently of whether or not they fall within the ambit of the Causal Role Problem: for example, 'the property of a group that it is cohesive; of a mental state that it is the belief that *p*; of a biological trait that it maximizes inclusive fitness' (1990: 112). For another, Jackson and Pettit do not seem fully committed to the claim that a merely 'causally relevant' (as opposed to efficacious) property really is distinctively causal; indeed, at one point they describe such a property as 'perfectly inert' (1990, 114). So to my mind one can read Jackson and Pettit's position as an argument for the explanatory usefulness of functional properties, even assuming that they are genuinely epiphenomenal. Indeed, Jackson is an unashamed epiphenomenalist: 'functional properties do not do any causing', he says (2012: 278).

The second reason why epiphenomenalism is not as bad as it may seem at first blush is that it is entirely consistent with the causal efficacy of mental *events*. Take your having been in pain just now. According to the general Lewisian story described in §2, and indeed—as we saw in §3—according to the view explicitly endorsed by Lewis in the particular case of mental events, there is no event that is *essentially* your being in pain; however, that property is a perfectly good accidental property of some event whose essential properties are physical. That event it is a perfectly legitimate cause of your pain behaviour, and it is a mental event in the sense that it is correctly describable in mental terms. So mental events can – and very often do – cause things. When people ingest hallucinogens, their doing so really does frequently cause them to hallucinate, notwithstanding the fact that the events we're quantifying over are only accidentally ingestions of hallucinogens.

Thirdly and finally, it is not a consequence of epiphenomenalism that mental properties are useless for the purposes of manipulating and controlling ourselves or our environment. By way of a contrast, consider some pair of properties *P1* and *P2*, such that we are perfectly well able to identify and intervene on instances of both *P1* and *P2*, and such that each of *P1* and *P2* (but no other property—and we know this) is causally relevant to some other property *Q*. Now suppose that we define a new property, *P**, such that to have *P** is to have some property or other that typically causes *Q*. *P** is, it would seem, a pretty useless property for practical purposes (though the predicate ‘*P**’ might be a convenient linguistic shortcut in certain circumstances). Our conceiving of the world in terms that invoke *P** does not really *help* us to bring about or avoid *Q*, in the sense that we can perfectly well intervene on either *P1* or *P2* themselves in order to bring about or avoid *Q*. (And of course if we can’t intervene on *P1* or *P2* for some reason, we aren’t going to be able to intervene on *P** either.)

The practical situation with respect to mental properties, however, is different. By and large, we cannot (as things currently stand, at any rate) intervene on subvenient physical properties directly, and in at least many cases it would probably be immoral to do so even if we could. Our only option is to intervene on their supervening mental cousins. Sally currently believes that the train doesn’t leave for another half hour, and her being in the physical state that realizes this mental property is causing her to get ready too slowly. *I* know that the train leaves in twenty minutes, and, since I want her not to miss it, I need to cause her to be in a physical state that realizes *that* mental property. I am able to do this in a variety of ways that, fortunately, require no knowledge of the physical state of Sally’s brain beyond which relevant mental states it is realizing: I point at the clock, show her the train timetable, or whatever. In other words, I intervene on her physical state *by* intervening on her mental state (or rather, strictly speaking, by doing something that, for all practical purposes, is just *like* intervening, intervention being a causal notion, and mental properties being just as unsuited to the role of effect as they are to the role of cause. This is, of course, a very common pattern in our manipulation and control of ourselves and others.

Nothing in this story requires us to conceive of mental properties themselves as bearing causal relevance to the behaviour in question. If I urgently need a writing implement—say because I’ve just witnessed a hit-and-run and I’m in danger of forgetting the car’s registration number—I might reasonably request something to write with. I don’t care whether it’s a fountain pen or a biro or a pencil or You satisfy my request by handing me a pencil, and I successfully write down the car’s registration number. We can make sense of the fact that this episode comes to a satisfactory resolution without having to conceive *being a writing implement* as itself causally relevant to my writing down the registration number. Nor must we commit ourselves to the view that an event that has *passing me a writing implement* as an essential feature has occurred. Similarly, if epiphenomenalism is true, for the mental case. We need not conceive of mental properties as causally relevant, or as essential features of any events, in order to make sense of our manipulation (or perhaps ‘manipulation’) of them in achieving our ends.

My aim here has merely been to point out that epiphenomenalism—of the kind that is engendered by a commitment to non-reductive physicalism, or so I’ve argued—is not *obviously* a crazy position. Our common-sense theory of the mental plainly assigns explanatory status to mental properties (‘Jack ate the chocolate because he was hungry’) and causal efficacy to mental events (‘Jill’s thumping headache caused her to turn the lights down’). And it’s a plain fact of life that we routinely control our own and others’ mental states and ensuing behaviour by intervening (or at any rate by doing things that for practical purposes are just like intervening) on the mental properties we and others instantiate. Epiphenomenalism, I have argued, allows the nonreductive physicalist to keep all of this. So what, exactly, doesn’t it allow us to keep, that we really care or ought to care about? This, I think, is the question that those philosophers who reject epiphenomenalism out of hand need to answer.

Bibliography

Antony, L. 2008. ‘Multiple Realization: Keeping it Real’. In *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by J. Hohwy and J. Kallestrup, 164–75. Oxford: Oxford University Press.

Bennett, K. 2007. 'Mental Causation'. *Philosophy Compass* 2: 316–37.

Bird, A. 2007. *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.

Block, N. 1990. 'Can the Mind Change the World?'. In *Meaning and Method: Essays in Honor of Hilary Putnam*, edited by G. Boolos, 137–70. New York: Cambridge University Press.

Jackson, F. 2012. 'Leibniz's Law and the Philosophy of Mind'. *Proceedings of the Aristotelian Society* 112: 269–83.

Jackson, F. and P. Pettit, 1988. 'Functionalism and Broad Content'. *Mind* 97: 381–400.

Jackson, F. and P. Pettit, 1990. 'Program Explanation: A General Perspective', *Analysis* 50: 107–17.

Kim, J. 1989. 'The Myth of Nonreductive Materialism'. *Proceedings and Addresses of the American Philosophical Association* 63: 31–47.

Levin, J. 2013. 'Functionalism'. *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), edited by Edward N. Zalta , URL = <http://plato.stanford.edu/archives/fall2013/entries/functionalism/>.

Lewis, D. K. 1980. 'Mad Pain and Martian Pain'. In *Readings in the Philosophy of Psychology* Vol. I., edited by N. Block, 216–22. Cambridge, MA: Harvard University Press, 1980.

Lewis, D. K. 1986a. 'Events'. In his *Philosophical Papers, vol. II*, 262–9. Oxford: Blackwell.

Lewis, D. K. 1986b. 'Causal Explanation'. In his *Philosophical Papers, vol. II*, 212–40. Oxford: Blackwell.

List, C. and P. Menzies. 2009. 'Nonreductive Physicalism and the Limits of the Exclusion Principle'. *Journal of Philosophy* 106: 475–502.

Ludwig, K. 1994. 'Causal Relevance and Thought Content'. *The Philosophical Quarterly* 44: 334–53.

Lyons, J. C. 2006. 'In Defense of Epiphenomenalism'. *Philosophical Psychology* 19: 767–94.

McLaughlin, B. 2007. 'Mental Causation and Shoemaker-Realization'. *Erkenntnis* 67: 149–72.

Molnar, G. 2003. *Powers: A Study in Metaphysics*, edited by S. Mumford. Oxford: Clarendon Press.

Pettit, P., THIS VOLUME. 'The program model, difference-makers, and the Exclusion Problem'.

Prior, E. W., Pargetter, R., and F. Jackson. 1982. 'Three Theses about Dispositions'. *American Philosophical Quarterly* 19: 251–57.

Raatikainen, P. 2010. 'Causation, Exclusion, and the Special Sciences', *Erkenntnis* 73: 349–63.

Rupert, R. 2006. 'Functionalism, Mental Causation, and the Problem of Metaphysically Necessary Effects'. *Nous* 40: 256–83.

Segal, G. 2009. 'The Causal Inefficacy of Content'. *Mind and Language* 24: 80–102.

Shoemaker, S. 2001. 'Realization and Mental Causation'. In *Physicalism and its Discontents*, edited by C. Gillett and B. M. Loewer, 74–98. Cambridge: Cambridge University Press.

Weslake, B. forthcoming. 'Exclusion Excluded'. *International Journal for the Philosophy of Science*.

Woodward, J. 2008. 'Mental Causation and Neural Mechanisms'. In *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by J. Hohwy and J. Kallestrup, 218–62. Oxford: Oxford University Press.

-- THIS VOLUME. 'Intervening in the Exclusion Argument'.

Yablo, S. 1992. 'Mental Causation'. *The Philosophical Review* 101: 245–80.