

## 4

# Mental Causation: Ontology and Patterns of Variation

PAUL NOORDHOF

Physicalism was initially motivated by its ability to deal with the problems of mental–physical interaction. The most attractive version of physicalism, though, is one which allows the mental some degree of autonomy with regard to the physical. Few physicalists feel driven to defend the claim that mental properties are identical with those which are identified by some suitably refined version of current physics. Unfortunately, as is only too familiar, non-reductive physicalism—that which denies such an identification—seems to have significant problems with mental causation of its own. In this paper, I begin by setting out the challenge to its efficacy due to Jaegwon Kim. I shall do this briefly because I am sure the reader has, by now, tired of seeing this argument stated. I just want to make a couple of comments upon it for the discussion ahead. I then discuss two over-reactions to it—one which seeks to understand mental efficacy non-ontologically in terms of patterns of variation, the other of which uses the problem to motivate a particular ontology—trope metaphysics. I explain why I consider these over-reactions, identify what is unsatisfactory about them, and then take elements of each to motivate my own approach which, you may not be surprised to learn, captures what is best in both. It also deals with an issue about mental causation untouched by Kim’s initial challenge.

### I. Kim’s Exclusion Argument

I’m going to set out Kim’s argument making certain assumptions to fix ideas. These assumptions don’t change the import of the argument, nor affect the

responses to it I am going to consider. They just simplify presentation. Specifically, I will assume that the non-reductive physicalist is committed to holding that there is more than one arrangement ( $A_1, A_2, A_3 \dots$ ) of narrowly physical properties ( $P_1, P_2, P_3 \dots$ )—those properties identified by current physics or a future development of it which suitably resembles it—such that, for each of them, it is metaphysically necessary that if they are instantiated, then a certain broadly physical property (BP) is instantiated (henceforth I will use upper case letters to designate type-properties and lower case letters to designate specific instances). One subclass of broadly physical properties is that of mental properties and behavioural properties. It is because all properties instantiated in the world are either narrowly physical or broadly physical, that non-reductive physicalism is true.

The appeal to metaphysical necessity is required to capture the fact that non-reductive physicalism is committed to a tighter connection than mere nomological necessity between arrangements of physical properties and mental properties. The latter type of connection would be acceptable to the emergent dualist. Debate has raged over whether appeal to metaphysical necessity is sufficient to capture what is required. I have defended this conclusion (Noordhof 2003, 2010). Nevertheless, all that matters is that it is stronger than the relation allowed by emergent dualists (bracketing an issue I touch on in section 5 about a powers ontology). If it is not, not only do we not have a version of physicalism but the issue set aside in the comment below about other events in the causal chain or causal circumstances becomes salient.

Appeal to metaphysical necessitation may appear too strong (e.g., Kim 2005, 49). Consider the relationship between  $O$ , a property occupying a certain causal role  $R$ , and the property of having role  $R$ . On some accounts of the connection between properties and laws, the relationship between  $O$  and  $R$  is one of merely nomological necessity. Laws independent of  $O$ , but governing its causal relations, give  $O$  the  $R$ -role. Nevertheless, the thought runs, the instantiation of the property of having role  $R$  is *explained* by the presence of  $O$  and the laws which hold relating to  $O$ . Instead of metaphysical necessity, we have nomological necessity plus explanation. This issue can be set aside by allowing that narrowly physical laws—those identified by physics—can be part of the metaphysical necessitation-base for a property. Thus we do have metaphysical necessitation still in play between  $O$ s and laws on the one hand, and  $R$  on the other.

With these assumptions in place, the argument against the efficacy of those broadly physical properties recognized by non-reductive physicalism runs as follows.

- (1)  $A_1(p_1, p_2, p_3 \dots)$  is causally sufficient for, or fixes the probability of,  $A_2(p_{100}, p_{101}, p_{102} \dots)$  (necessitation-bases for, but not identical to,  $bp_1, bp_2$ , respectively).
- (2)  $bp_1$  is a cause of  $bp_2$  (Assumption).
- (3)  $bp_1$  causes  $bp_2$  either directly or by causing  $A_2(p_{100}, p_{101}, p_{102} \dots)$ .
- (4) If  $bp_1$  causes  $bp_2$  directly, then either  $A_1(p_1, p_2, p_3 \dots)$  is insufficient for  $bp_2$  by causing  $A_2(p_{100}, p_{101}, p_{102} \dots)$  or  $bp_1$  is an overdetermining cause.
- (5) If  $bp_1$  causes  $bp_2$  by causing  $A_2(p_{100}, p_{101}, p_{102} \dots)$ , then the same choice holds regarding  $A_2(p_{100}, p_{101}, p_{102} \dots)$ .
- (6) There is no systematic overdetermination in this way.

Therefore,

- (7)  $bp_1$  is inefficacious (see, e.g., Kim 1998, 41–7; Kim 2005, 39–52).

The argument does not claim that if  $A_1(p_1, p_2, p_3 \dots)$  is causally sufficient for  $A_2(p_{100}, p_{101}, p_{102} \dots)$  then there can be no other sufficient cause without overdetermination. There may be other sufficient causes which are part of the causal circumstances, or further up or down the causal chain. The focus is just on the efficacy of  $A_1(p_1, p_2, p_3 \dots)$  and  $bp_1$  for target effects  $A_2(p_{100}, p_{101}, p_{102} \dots)$  and  $bp_2$  standing in the same relationship. The question is whether, at that point in the causal network, there is any contribution for  $bp_1$  to make given  $A_1(p_1, p_2, p_3 \dots)$ 's presence.

Talk of position in a causal network may raise alarm bells because of putative difficulties in fitting conditions, in particular negative conditions, into the framework (Steward 1997, 135–40). It should not. The argument does not require an exhaustive causal network. All that is required is that we can make sense of the idea that token events, or property instances, stand in a causal network against a backdrop of assumed causal conditions and that, as a result of this, we can see two or more events, or property instances, as in potential competition for efficacy at a certain position in this network.

Although I have dubbed this argument Kim's exclusion argument, the appeal to the causal exclusion principle is implicit. The principle holds that

No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination (Kim 2005, 42).

(4) and (5) each claim that the choices are insufficiency of one of the putative causes or overdetermination. This is what the causal exclusion principle claims. I do not appeal to the causal exclusion principle explicitly because it is inadequately formulated given the first point I made about what the argument does not claim. At a given time, there may be two or more sufficient causes each of which is sufficient, given causal circumstances that include the other of the causes.

The argument involves a simplification relating to Jaegwon Kim’s distinction between supervening and micro-based properties. My appeal to metaphysical necessitation does not distinguish between these two cases. Nevertheless, Kim holds that the kind of argument I rehearse works against the former and not the latter. Since the argument only seems to need to appeal, at the crucial point, to the idea that  $A_1(p_1, p_2, p_3 \dots)$  is sufficient for, or fixes the probability of,  $bp_2$  by being sufficient for, or fixing the probability of, something which is sufficient for  $bp_2$ , it is hard to see how to justify the distinction between the cases (for more detailed discussion, see Kim 1999; Noordhof 1999b, 2010).

The argument also works at a certain level of abstraction that may seem to reduce its threat or make its application uncertain. Candidate BPs will include those we attribute by attributing the belief that ... where ‘...’ is filled in by some specification of content, sensation of ... where ‘...’ might be filled in by ‘burning feeling in the foot’, and so on. Those who put forward the argument, and those who discuss it, often work with the standard picture that  $A_i(P_j, P_{j+1}, P_{j+2} \dots)$  refers to some arrangement of narrowly physical properties in a subject’s brain. It may well be plausible that the following is true:

If S has  $BP_1$  and  $BP_2$ , and  $bp_1$  causes  $bp_2$  (where these are the particular instances of the properties attributed to S), then there are some arrangements of narrowly physical properties in S’s brain, say  $A_1(p_1, p_2, p_3 \dots)$  and  $A_2(p_{100}, p_{101}, p_{102} \dots)$ , which are *part of* the metaphysical necessitation-bases of  $BP_1$  and  $BP_2$ , and  $A_1(p_1, p_2, p_3 \dots)$  causes  $A_2(p_{100}, p_{101}, p_{102} \dots)$ .

That is, corresponding to mental efficacy, there is related efficacy at the narrowly physical level in the brain. However, it is no part of the argument that this assumption is written in. All it needs is the idea that, however

extensive the metaphysical necessitation-base for mental properties needs to be, putative causal relations between them imply corresponding causal relations between these bases (where a metaphysical necessitation-base for a property is one whose instantiation metaphysically necessitates the instantiation of the property in question).

Two reactions to Kim's argument are popular but I will argue are over-reactions. The first says that, in fact,  $bp_1$  is efficacious because it stands in different patterns of variation to  $bp_2$  than  $A_1(p_1, p_2, p_3 \dots)$  as a result of which it plays a distinctive causal explanatory role. I say that this relies on an understanding of causation that is, at once, too strong and too weak. We don't have to adopt an account of causation with such counterintuitive consequences (as we shall see) to have an answer to the exclusion argument. We should not take distinct patterns of variation, and the inferential consequences which flow from this, as fully capturing the reality of causation. The second says that  $bp_1$  is efficacious because it is identical to  $A_1(p_1, p_2, p_3 \dots)$  even though  $BP_1$  is not identical to  $A_1(P_1, P_2, P_3 \dots)$ —we have a property instance identity without an identity of properties. One version of the latter proposal—attractive because it provides a prima facie answer to an immediate objection—is formulated in terms of a trope metaphysics. Here I will argue that instance identification is, in itself, questionable, inadequate to support the whole weight of the response and leads one to a dubious metaphysics. Most importantly, it conflates property causation with property instance causation in its attempt to provide a defensible position. I consider these responses in turn in the next two sections as preliminaries to my own preferred approach, with, of course, nary a hint of over-reaction to be found.

## 2. Different Patterns of Variation

The first line of response to the argument appeals to, in the limiting case, different patterns of absence. For example, it is noted that the following are true.

(PA1) If  $bp_1$  had not occurred, then no necessitation-base of  $bp_1$  would have occurred.

Hence there would be no  $bp_2$ .

(PA2) If  $A_1(p_1, p_2, p_3 \dots)$  had not occurred, then another necessitation-base of  $bp_1$  might have.

Hence, it is not the case that there would be no  $bp_2$  (e.g. List and Menzies 2009, 487–9; Menzies 2008, 210).

Truths such as this have been used in various contexts. Sometimes it is said that causation is a contrastive matter. The basic form of causation is that  $c$  rather than  $c'$  causes  $e$  rather than  $e'$ . Different ways of describing what might be thought to be one property instance (e.g.,  $A_1(p_1, p_2, p_3 \dots)/bp_1$ ), or identifications of distinct property instances, set up different comparison classes. Contrastive accounts of causation typically take property instances (or events) to be coarsely individuated—so that a property instance involves the co-instantiation of multiple properties—because the case for more finely individuated property instances (or events) can be answered if causation is contrastive (e.g., Schaffer 2005, 347). Nevertheless, this is not mandatory.

Taking causation to be contrastive is often accompanied by the claim that causal statements are context-sensitive. Context-sensitive statements convey different propositions in different contexts of use. In the case of causation, the context-sensitivity concerns what is the *foil* to the target property instance or event. If there is variation in this—because, in some contexts, the foil is *absence* of the target event, in others a specific *alternative* event—then causal statements would be context-sensitive in this respect.

(PA1) and (PA2) take the foil to be the absence of a property instance satisfying a particular description. Thus, describing a property instance as  $bp_1$  determines the comparison class to be the absence of any property instance correctly described as an instance of  $BP_1$ . This will include lots of other necessitation-bases of  $BP_1$ . Whereas, describing a property instance as  $A_1(p_1, p_2, p_3 \dots)$  determines the comparison class to be the absence of any property instance correctly described as an instance of  $A_1(p_1, p_2, p_3 \dots)$  (see, e.g., Menzies 2008, 206–8). By focusing simply on the case of absence, as (PA1) and (PA2) do, we bracket the question of context-sensitivity. Nevertheless, the considerations offered below with regard to the case of absence may be generalized.

Counterfactual theories of causation promise an immediate explanation of the relevance of (PA1) and (PA2) though, as we shall see, this promise is not kept. They are generally formulated as contrastive theories in which the contrast is always with the absence of the target cause. Non-counterfactual

theories have to generate (PA<sub>1</sub>) and (PA<sub>2</sub>) either from taking their truth as a constraint—their approach is geared to make such counterfactuals true—or by writing in the contrastive component as an additional element, for example, by holding that in a layered world of natural kinds ‘same level causation is the norm’ (Gibbons 2006, 88, where the talk is of systematic difference-making rather than explicitly of counterfactuals such as these which express difference-making).

In any event, the claim is that the mental is shown to be efficacious, by identifying the right difference-making as plausibly revealed in such counterfactuals. The reasoning runs as follows. ‘If *c* were not the case, then *e* would not be the case’ is a plausible sufficient condition for causation.  $A_1(p_1, p_2, p_3 \dots)$  fails to satisfy this condition if (PA<sub>2</sub>) is true. That is, if there might be some other necessitation-base of  $bp_1$  so that  $bp_2$  may still be the case. By itself, this doesn’t show that  $A_1(p_1, p_2, p_3 \dots)$  fails to be a cause if the counterfactual dependence of *e* upon *c* were merely a sufficient condition. So it seems that it is being taken as a necessary condition too, in the circumstances.

The first thing to note is that we don’t allow that the possible occurrence of replacements to a cause to discredit that cause from being efficacious, on pain of making the world’s causal processes very gappy affairs. For example, suppose my head of department comes to me and points out that I have done very little administration for the department recently and other folks have done lots of stuff. I don’t undermine what they have done by saying that, since, if they hadn’t done it, I would have done it in their place, they cannot be credited with having done anything. Yet, the situation seems analogous. There were two, or doubtless more, possible undertakers of these administrative tasks. Undertaking these tasks was just realized in them rather than me. Causes are those things which are actually involved in the process which led to a certain target effect. Otherwise, at every point in the process at which there might have been a replacement, we would have a causal gap. Of course, you could decide to call the gap ‘a gap of causation’—a gap which is filled by the occurrence of actual determination—but the decision to talk this way has no particular utility and, as we shall shortly see, would not alleviate the worries about mental causation in any case.

A second, and related point is that the counterfactual reasoning which is meant to support  $bp_1$ ’s efficacy over  $A_1(p_1, p_2, p_3 \dots)$ ’s claims is the same reasoning that is judged inappropriate in all cases of redundant causation,

especially pre-emption. Recognizing the existence of pre-emptive causation precisely turns on supposing that the possible occurrence of replacements does not undermine the pre-empting cause's entitlement to be called a cause. It is a significant cost to appeal to a pattern of reasoning which would discredit all pre-emptive causes in order to discredit the causal claims of  $A_1(p_1, p_2, p_3 \dots)$ , and yet this is what is required to answer the exclusion argument without rejecting a version of the exclusion principle. Perhaps Peter Menzies will say that it is appropriate to appeal to this pattern of reasoning given that redundant causation has been tacitly ruled out in this kind of case. However, it is hard to see this move as legitimate bearing in mind that no characterization of the difference between this type of case and redundant causation has been provided and the latter is very much an option which is under consideration in discussions of this issue.

An unfortunate consequence of the patterns of variation approach is that it makes the resolution of Kim's argument turn upon brain plasticity, in our terminology, the plasticity of arrangements of narrowly physical properties supporting the causal relationship between mental properties. If it is the case that no replacement arrangements of narrowly physical properties would subserve the relationship, if the actual arrangement of physical properties were absent, then we would be back with causal competition once more with mental properties the potential losers. This may not, in fact, be an issue because neuroscientists have observed that, as a result of damage, different parts of the brain can be used to play the same function. Nevertheless, it would be surprising if the efficacy of mental properties turned on whether or not this held on a case by case basis. Furthermore, since brain plasticity reduces with age, this proposed response seems stuck with the potential consequence that subjects' mental properties may lose efficacy during the course of their lives.

The counterfactual reasoning with which I began this section has been taken to express another feature of causes which, thereby, provides a motivation for taking the previous points I've made to be inconclusive. This is the idea that causes should be proportional to their effects and not contain lots of redundant elements (Yablo 1992; Menzies 2008; List and Menzies 2009, 488–9). This is alleged to be the difference between  $bp_1$  and any of the  $A(\dots)$ s.

As things stand, this last claim is susceptible to a deflationary response. The objector to the efficacy of  $bp_1$  can concede that talk of  $bp_1$  has a causal implication that talk of  $A_1(p_1, p_2, p_3 \dots)$  does not:  $bp_1$ 's absence ensures the



absence of sufficient causes/chance-fixers, that is, any of the  $A(\dots)$ ; talk of a particular  $A(\dots)$  does not. Nevertheless, it can be argued, it is not that, by these means,  $bp_1$  is revealed to be the cause itself.  $bp_1$  is not, by anybody's lights, a cause of any of its necessitation-bases, rather its absence *entails* the absence of any of them. In brief, we have *causal explanatory* impact without causation.

The claim of proportionality is, plausibly, overstated in any case and, thus, doesn't get past the difficulty raised by the exclusion argument. Considerations of proportionality entitle something to be counted a cause in the following sense.

If  $c$  and  $c'$  are putative competitor causes of  $e$  at the same point in the causal network, and  $c$  is more proportional than  $c'$  for  $e$ , then if  $c'$  is a cause,  $c$  is a cause.

In brief, the reason for this is that more proportional causes are specified in terms of properties which enable us to capture a generality that less proportional causes miss. So if the latter is a cause, the former will be too. This will become clearer on the development of my own approach in section 5. For the moment I observe, first, that the attempts to discredit the efficacy of  $A_1(p_1, p_2, p_3 \dots)$  have not been successful and, second, there is no motivation for adopting a distinction between causing on the one hand, and causal sufficiency for, or determination of, the probability of the target effect on the other. Causing something may involve additional features than simply being causally sufficient for, or a determinant of, the probability of the target effect. Nevertheless, the latter is plausibly a necessary condition for the former and, as a result, an exclusion argument run in terms of causal sufficiency or determination would appear almost as damaging, if not as damaging (for more discussion, see Noordhof 1999c, 374–5).

Thus, we are left with an apparent causal explanatory difference that we must evaluate to see whether we have a corresponding difference in causal reality. To conclude that difference in causal reality just falls out of the causal explanatory difference identified is the first of the two over-reactions I promised to identify.

### 3. Identity of Property Instances

An alternative fashionable approach to Kim's argument is to argue that  $bp_1$  is identical to  $A_1(p_1, p_2, p_3 \dots)$ . Although the properties,  $BP_1$  and  $A_1(P_1, P_2,$

$P_3 \dots$ ) are not identical, when it comes the instances— $bp_1$  and  $A_1(p_1, p_2, p_3 \dots)$ —they are. It cannot be denied that, *if* these property instances are identical, then this particular problem is resolved.  $bp_1$  is a cause given agreement that  $A_1(p_1, p_2, p_3 \dots)$  is, and so on for all other instances of broadly physical properties so identical. Evaluation of this approach doesn't focus on whether it works so much as whether, and indeed how, the identification can be justified. On the positive side, there is the satisfactory result for the problem of mental causation the exclusion argument raises. But are there things to be said on the negative side? Identifications need to be justified by more than the fact that they offer a convenient simplification of our problems. They need to be independently plausible or, at least, not implausible.

Some will take this challenge as unfair. They will remark that identity is a primitive relation so one cannot expect any justification of it. To the extent that we need a reason for recognizing the identity, resolution of the problem of mental causation in this vicinity supplies us with one. However, this is mistake. First, there is the slide from metaphysics to epistemology. Identity may be a primitive relation but that doesn't mean that justification for supposing it to hold must be taken to be primitive. We may have complex reasons to believe simple things. Second, the combination of views proposed is that two properties may be distinct yet have identical instances. We need an account of why this combination is coherent. Third, within a metaphysical framework which makes this combination of views possible, the considerations in favour of taking instances of mental properties to be identical with instances of arrangements of physical properties must have general application. We can't have mental property instances as a special case. That would be unmotivated.

It is no surprise, then, that sophisticated proponents of this strategy address these issues. It is convenient to divide the approaches into those which take properties as universals to be the fundamental element and those which take property instances or tropes to be the fundamental element. I shall consider these in turn.

The apparent problem for the first approach—which takes properties as universals to be the fundamental element—is how one instance could involve the instantiation of two distinct fundamental elements. Instance identity and distinctness, it would seem, must follow universal identity and distinctness (Ehring 1997, 462–3). The following makes the connection explicit.

An instance of *F* is identical to an instance of *G* only if  $F = G$  (where ‘*F*’, ‘*G*’ are universals).

A sufficient condition for instance identity will draw upon additional factors that serve to distinguish between instances, e.g. spatiotemporal location.

This problem seems overstated. The, by now, standard, subset, approach to property co-instantiation seems available to those who take properties to be universals. The subset view of property instance identity holds that  $bp_1$  is identical with  $A_1(p_1, p_2, p_3 \dots)$  if the causal powers of  $BP_1$  are a subset of the causal powers of  $A_1(P_1, P_2, P_3 \dots)$ . A property instance with a set of causal powers  $\{CP_1, CP_2, CP_3, CP_4 \dots\}$  will count as a property instance of a property with that set, and also of a property with, say,  $\{CP_1, CP_2 \dots\}$  alone (Whittle 2007, 68–9, who doubts that the subset view can be used in the straightforward fashion recommended here). The apparent distinctness of the instantiation of *F* and the instantiation of *G* is shown to be mistaken because the causal powers of the former are a subset of the causal powers of the latter. Co-instantiation as partial coincidence in causal powers is not ruled out to those with universals in their ontology.

Some prefer to say that  $bp_1$  is realized by, but is not identical to,  $A_1(p_1, p_2, p_3 \dots)$  when the causal powers of the former are a subset of the causal powers of the latter (Shoemaker 2007, 17; for other grounds for resisting identity, see *ibid.* 48–9). They rightly point out that instances cannot be identical if the causal powers of one stands in the subset relation to the causal powers of the other (or, with qualifications, more precisely, causal profile which includes the ways in which an instantiation may be caused too, as well as its causal powers, see Shoemaker 2007, 11–12, 16–17). Their strategy is not strictly speaking an example of the identity of instance strategy; however, it resembles it in important respects. They take it that there is a state of affairs  $A_1(p_1, p_2, p_3 \dots)$  which realizes  $bp_1$ , by having its causal powers as a subset, and  $bp_1$  is efficacious when the subset of causal powers relating to it are in play. Talk of states of affairs allows them to resist characterizing  $A_1(p_1, p_2, p_3 \dots)$  itself as a property instance—as opposed to an arrangement of property instances—while retaining the subset picture. They can, of course, allow that there is a property of being a certain kind of state of affairs if they wish (e.g., Shoemaker 2007, 32–4). The difference between instance-identity and realization just mentioned brings out the slipperiness of the term ‘co-instantiation’. It can either mean identity of

instances or coincidence of instances in a particular instantiation. My remarks below apply equally to both characterizations.

The basic problem is that the subset view is not true if you take the existence of BP properties seriously. To fix ideas, consider a case of pain. A part of me hurts and my experience of it—my pain experience—is of a type that causes writhing and general unhappiness in humans and, we may presume, similar writhing and unhappiness in sentient robots. The hallmark of this activity is that the writhing is because of heat damage done to an arm. So the writhing stems from there and is, in some way, directed towards alleviating what is going on there.

In humans, this property is necessitated by particular kind of  $A\delta$ -fibre or C-fibre firing (let us suppose), which kind depending upon the type of hurting involved. Let's focus on the famous C-fibre firing and take the particular hurting to be necessitated by C-fibre firing in way W. In robots, the pain will be necessitated by something different, let's call that C-circuit activity in way V. Now the question is whether the causal powers of pain are a subset of C-fibre firings' causal powers and of C-circuit activity's causal powers. The answer seems to be no. Pain experiences have the capacity to cause pain behaviour in humans and robots whereas C-fibre firing in way W can only cause such pain-behaviour in humans, and C-circuit activity in way V can only cause such pain-behaviour in robots. So, the subset view would deny that instances of pain experience of the kind specified are identical with, or for that matter realized by, either.

One move would be to claim that pain-experiences-in-humans and pain-experiences-in-robots, rather than simply pain experiences, are to be identified with C-fibre firings and C-circuit activations. This was Kim's proposed response to the existence of variable realization in defence of reductive physicalism (i.e., type-type identity theory) (Kim 1992, 330–5). It rested upon the claim that the nomic relationships, and hence causal powers, of instances of mental properties or states are to be explained in terms of the nomic relationships, and causal powers, of narrowly physical properties (Kim 1992, 322). As we shall see later, this claim is susceptible to a number of different interpretations. Its use in this context, though, is questionable. Denying that there are pain experiences, as opposed to species-specific pain experiences, is, quite obviously, refusing to take them seriously.

The existence of a cross-species psychology reflecting general truths about the causal implications of having pain experiences requires more

than merely species-specific pain states. Information about one type of species-specific pain experience would, if pain experiences failed to exist, imply nothing about the role of pain experience in other creatures. Yet, when we reflect upon how a creature would respond if something hurt like this—thinking about a particular kind of pain experience—we think that there are general psychological commonalities between how they would respond, and how we would respond, independent of variation of physical constitution. There may be differences too, as a result of our differences in physical constitution, but recognizing that there may be psychological differences is quite compatible with also recognizing commonalities. This is a phenomenon with which we are familiar for individuals too. Recognizing individual differences does not imply that all that is possible are individual psychologies and individual-relative states.

Let me state a bit more precisely how this might work in the face of Jaegwon Kim's scepticism (e.g., Kim 1992, 323–5). Suppose that BP1 is metaphysically necessitated by each of the following arrangements of narrowly physical properties  $A_1(P_1, P_2, P_3 \dots)$ ,  $A_1(P_{11}, P_{21}, P_{31} \dots)$ ,  $A_1(P_{12}, P_{22}, P_{32} \dots)$  ... Then, BP1 can be related to radically disjunctive narrowly physical conditions— $A_1(P_1, P_2, P_3 \dots)$ ,  $A_1(P_{11}, P_{21}, P_{31} \dots)$ ,  $A_1(P_{12}, P_{22}, P_{32} \dots)$  ...—in that the corresponding causal powers of these conditions are  $\{CP_1, CP_2, CP_3, CP_4, CP_5, CP_6 \dots\}$ ,  $\{CP_1, CP_2, CP_{31}, CP_{41}, CP_{51}, CP_{61} \dots\}$ ,  $\{CP_1, CP_2, CP_{32}, CP_{42}, CP_{52}, CP_{62} \dots\}$ . The vast majority of the causal powers are disparate.

According to the subset approach, instances of  $A_1(P_1, P_2, P_3 \dots)$ ,  $A_1(P_{11}, P_{21}, P_{31} \dots)$ ,  $A_1(P_{12}, P_{22}, P_{32} \dots)$  ... are each instances of BP1 because a subset of their causal powers are  $\{CP_1, CP_2\}$ , the causal powers of BP1. Some of the other causal powers listed may be responsible for psychological differences but there is a significant psychological commonality. The problem I'm raising for this approach with regard to the case of c-fibre and c-circuit activity is that the arrangements of narrowly physical properties that we find it plausible to count as instances, or realizations, of mental properties do not seem to have the causal powers that would constitute the commonality. That is,  $\{CP_1, CP_2\}$  are not part of the sets of powers, contrary to how they have been represented above.

A second move in response to this case may seem to help with the difficulty just identified: the distinction between core realization and total realization. The core realization of a BP is that arrangement of narrowly

physical properties which play the causal role associated with BP's instantiation. The total realization is the arrangement of narrowly physical properties *together with the context in which they occur* which, taken together, necessitate that an instance of BP is presently typically playing the role in question. As Sydney Shoemaker puts it, to motivate the distinction, firing C-fibres in a Petri dish is not a case of pain (Shoemaker 2007, 21; the distinction goes back to Shoemaker 1981). For mental properties to be instantiated, it seems plausible that not only must properties with a certain causal profile be instantiated but, in addition, key elements of the profile must be typically manifested.

With this distinction in place, it might be argued that, if C-fibre firing in way  $W$  occurs in a robot, then it fails to cause writhing because the total realization of the relevant experience of pain is not present. So an instance of BP does not display a causal power that an instance of C-fibre firing in way  $W$  fails to have. This response is mistaken. For BP to have a causal power  $A_1(P_1, P_2, P_3 \dots)$  fails to have, BP does not have to display that causal power in circumstances in which  $A_1(P_1, P_2, P_3 \dots)$  is present and does not display it. BP can display the causal power when it is realized in a different way and causes something that  $A_1(P_1, P_2, P_3 \dots)$  would not in those circumstances.

Appeal to the idea of total realization explains how individual differences are compatible with cross-species psychological laws. The total realizations of mental properties ensure that certain causal relations typically hold across differences of constitution. Compatible with this, other differences in constitution may result in differences of psychology which disrupt these typical causal relations. Psychological laws may have written into them conditions under the typical causal relations don't hold and these conditions may imply that the regularities do not hold at all in some other species. This is no more exceptionable than our appreciation that, for example, if we were more secure, a rejection would have less of a significant effect than it does in our case.

The distinction between core and total realization raises a question mark over the efficacy of instances of BP. If the conditions under which BP is necessitated include the circumstances, then can it be attributed causal powers with regard to those circumstances? An answer to this will come in the development of my own proposal in section 5. However, in brief, the response is that BP is efficacious in virtue of the fact that a part of its realization is efficacious.

A third move in support of the subset approach to mental causation is to claim that the causal powers of an instance of pain experience are no more than that of the property with which it is co-instantiated even though the property of being a pain experience has causal powers which exceed those properties with which it is co-instantiated. Two ways in which this might be achieved are either indirectly from a claim about the individuation of instances or directly from a claim about what is required for instantiation.

The indirect method would be to say that, suppose that a particular pain experience is co-instantiated with an instance of C-fibres firing in way W, then that instance of pain experience could not be co-instantiated with something else. There is no possible world in which, say, my instance of pain experience could be co-instantiated with C-circuit activity in way V. So, this particular instance of pain experience cannot have causal powers which exceed those of that with which it is co-instantiated.

Obviously, if the instance of pain experience is identical with an instance of C-fibre firing in way W, then it is plausible that they share modal properties. However, it would be illegitimate to appeal to instance identity by itself to establish that the causal powers of one of the instances don't outstrip those of the other. That's supposed to be the conclusion of the subset approach. Instance identity is not meant to establish the correctness of the subset approach. Furthermore, an actual future case, rather than appeal to a possible case, might put this under pressure. Suppose that I have a throbbing pain and I receive prosthetic neural fibre replacement without anaesthetic in the hope of stopping the pain and the pain persists. It is very plausible to say that that instance of pain experience—and not just a pain experience of the same type—is continuing although realized by different neural fibres.

It could be argued that the instances of prosthetic C-fibres have the same causal powers as the instance of pain experience in the case described. So they present no problem. By contrast, the instance of my pain experience could not be co-instantiated with C-circuit activity because I could not be a robot. The object in which a property is instantiated is, the claim would run, one of the essential features of an instantiation.

The success of this response partly turns on whether it is plausible to insist that I could not have been constituted from the same material as a robot. If I am essentially a particular human animal, then the response receives support. Those friendly to a psychological characterization of personal

identity will resist the claim that I could not be made from the same material as a robot and even those who insist the psychological view is false, and that we are animals, don't have to conclude that we are animals essentially (Olson 1997, 125, for the claim I could not become a robot; for the claim that I am an animal is compatible with my becoming a robot, see Olson 2007, 27). If I might be a robot, then it is possible that my life could have run a different course so that my pain experience now, in fact co-instantiated with C-fibre firing in way W, could have been co-instantiated with C-circuit activity firing in way V. In which case, the causal powers of the instance of pain experience threaten to outstrip that which is co-instantiated with it. So the instance of my pain experience is not identical with instances of either C-fibre firing or C-circuit activity.

Nor is this the extent of the problems the response faces. Certain cases of causation seem to involve the transfer of a property instance from one object to another. For example, there is a difference between an object being sticky on one side, it ceasing to be, and another object being sticky on one side, and the stickiness of the first object wholly transferring to the second object by contact. In the latter case, it is plausible to say that the same instance of the property of being sticky has moved from one object to the other. In which case, the object in which the property is instantiated cannot be essential to the identity of the instance (Ehring 1997, 123–4). It is usually thought that only a trope metaphysics can account for this kind of fact but this does not seem correct.

The reasoning goes like this. Exemplifications are essentially momentary. For suppose otherwise, then we should allow the same in the case of spatial location. If an exemplification of squareness on one side of my office is not distinct from an exemplification of squareness on the other side of my office, then we have an exemplification which is wholly present in two places. We have lost the distinction between exemplifications and universals. We would do the same if we recognized non-momentary exemplifications (Ehring 1997, 87–9).

This line of reasoning can be resisted. Exemplifications do not have to be momentary. They can have duration. Thus if an object remains square over a period of two hours, 11 a.m. to 1 p.m., there is a single exemplification of the universal of squareness. What should we say about the object's squareness at noon? Rather than recognize momentary exemplifications of squareness which, then, have to constitute the exemplification of squareness over



two hours, we should deny that there is an *exemplification of squareness at 12 noon*. There is an exemplification of squareness between 12 and 2 in virtue of which it is true that the object is square at 12 noon. Corresponding to every timely predicate, there does not have to be an exemplification.

The same applies to spatial extent. An expanse of red involves a single exemplification of red. We do not have to suppose that exemplifications of red at much smaller regions make up this single exemplification otherwise, amongst other difficulties, we would need to consider the smallest extent in which red could be realized and question whether that smallest extent could be described as red at various points within it. If it is allowed that it can be so described without there being an exemplification of red at the points with the smallest extent, then we might as well accept that there is a single exemplification of red across the extent with it still being true that portions of that single exemplification are also described as red.

The difference between what it is plausible to say in this case, and what I said previously concerning two spatially separated exemplifications of squareness, derives from the fact that the latter property has defined boundaries. So an exemplification of squareness falling outside the boundaries cannot constitute the same exemplification of squareness. Whereas, in the case of an instance of redness, the extent is not fixed. So we don't have to concede that a lesser extent must also count as an exemplification of redness. The claim that that lesser extent is red can be true simply in virtue of the larger extent which is an exemplification of redness.

Since exemplifications of properties are not individuated by the objects which possess them, there is nothing to rule out an instantiation of pain in me being instantiated in another creature with a different constitution. Even if they were so individuated, it is an additional step to hold that such individuation of exemplifications requires that the means of individuation is in terms of essential properties. We individuate objects by their spatio-temporal position. From that, it does not follow that their actual spatio-temporal position is essential to them.

The second version of the third move I identified in support of the subset approach was to tinker with what is required for instantiation. So it may be suggested that  $A_1(p_1, p_2, p_3 \dots)$  is identical to  $bp_1$  by having truncated causal powers of  $BP_1$ . This will still not take the existence of  $BP_1$  seriously unless you also allow that the same would hold for  $A_1(P_1, P_2, P_3 \dots)$ . Thus, an instance of the latter may be identical to  $A_1(p_1, p_3, p_4 \dots)$  because, whatever

causal powers it loses from not having  $p_2$  as part of the arrangement are not sufficient to undermine the instance identity. You can't resist this by claiming that  $p_2$  must be part of the instantiation of  $A_1(P_1, P_2, P_3 \dots)$  because instance identity is meant to be determined by the subset of causal powers relation—or now the truncated subset relation—and not what properties constitute the property to be instantiated. After all, if the latter were in play, we would have grounds for denying that instances of pain were identical with instances of arrangements of narrowly physical properties and the argument would centre around providing a justification for thinking that narrowly physical properties were constituents of pain. However, since it is unacceptable to hold that  $A_1(p_1, p_3, p_4 \dots)$  is an instance of  $A_1(P_1, P_2, P_3 \dots)$ —otherwise how would we distinguish between instances of  $A_1(P_1, P_2, P_3 \dots)$  and instances of  $A_1(P_1, P_3, P_4 \dots)$ —this must be because having a truncated set of causal powers is not enough.

The second problem with the truncated powers proposal is that our evidence about  $bp_1$  in us would provide no grounds for supposing that pain in robots would make silicon creatures writhe. The only causal powers that my pain would reveal to me would concern how it affected me. I would not be able reasonably to assert that that pain—in me—is so bad that, if it were instantiated in a robot, they would be writhing about too unless they had much greater powers of pain control to me and could focus their attention away from it. I'd have no idea at all—by the truncated powers view—what powers pain would have in robots. But this is not correct. The reason why we know how others would behave if they had pain—even if they had a different constitution—is that we know the causal powers that pain would have in them from our own case (Gibbons 2006, 95–7, also emphasizes this).

Our starting position was that there were two distinct properties understood to be universals—being a pain experience of a certain type and being c-fibre firing in way  $W$ —which, therefore, would naturally be thought to have distinct instances. We were looking for a justification for concluding that some of their instances are identical. The subset proposal fails to provide it. This may not be altogether surprising. Proponents of trope metaphysics take the situation to be different within their framework. In fact, many of the points travel across and trope metaphysics has problems of its own with providing what is needed.

Proponents of a trope metaphysics take property instances as fundamental and construct physical and mental properties from these elements. This

gives them a motivated way of avoiding the question of how distinct properties could be identical in instances because this will be built into the construction. In addition, by insisting that property instances are fundamental, they have a way of resisting the claim that they have a structure which will raise, once more, the question of whether it is the mental or physical component of the instance which is causally relevant. Thus, a response to Kim's argument would then be a full response to the question of causal relevance of properties. A further question cannot be raised about in virtue of what features of an instance, is that instance efficacious. Unfortunately, a substantial motivation in favour of their metaphysics tells against this solution to the problem of mental causation.

Trope metaphysicians construct properties from exact resemblance classes of tropes. This only works if two properties don't share the same instance. If two properties do share the same instance, and are not coextensive, then we cannot appeal to exact resemblance. We must appeal to rough resemblance and, indeed, that is what those who hold that mental properties share property instances with physical properties do.

Just in case, the issue isn't obvious, let me explain why they need to move to rough resemblance. Let  $m_1, m_2, m_3 \dots m_n$  be a particular class of mental tropes (e.g., each of which is a pain experience) and  $p_1, p_2, p_3 \dots p_n$  be a class of physical tropes. Remember that the physical class and the mental class cannot be coextensive because, according to non-reductive physicalism, mental properties are not identical to physical properties. Let  $p_3 = m_3$ . Then if the classes of M and P were constructed from exact resemblance, then  $m_3$  would exactly resemble all the  $m$ s and exactly resemble all the  $p$ s too. But in that case, each of the  $p$ s would exactly resemble each of the  $m$ s (because exact resemblance is transitive) and we would just have one class after all (Gibb 2004, 471–2). Or, put it another way,  $m_1$  and  $m_2$  must exactly resemble each other. But if they are identical to different physical properties (that is, if they are variable realized), then those physical properties cannot exactly resemble each other. Again, to be able to construct a mental property class, trope metaphysicians need to appeal to rough resemblance. Members of a class roughly resemble each other to a certain degree which is greater than any non-member.

The appeal to rough, rather than exact, resemblance undermines the motivation for trope metaphysics in the first place. A principal reason for adopting a trope metaphysics, rather than resemblance nominalism, is

because of the problem of imperfect community (Campbell 1990, 32–4, 72–3). The difficulty identified for resemblance nominalism under this heading is that if you try to construct universals from resemblance classes of objects, then you will be committed to surrogates for universals which are not united in *a* resemblance but rather may be united by different resemblances between objects in the specified class. Moreover, it is unclear that even some of these classes will be surrogates for universals because a class bound together by a resemblance may be legitimately made a more extensive class in virtue of other resemblances if all that is required is rough resemblance (Manley 2002, 77–9, who raises the latter difficulty for worlds with limited members but it is not clear that more populous worlds avoid the difficulty).

Trope metaphysicians claimed to avoid the problem of imperfect community by appealing to the notion of exact resemblance as opposed, simply, to resemblance. Appeal to exact resemblance only works, though, for the trope metaphysician, if each property instance is a property instance of only one property (Campbell 1981, 134–5; Campbell 1990, 66, 72–3; abandoning his Campbell 1981 position on p. 137). Once you allow that a property instance may be a property instance of two or more properties the problem reasserts itself.

The result is that proponents to the trope metaphysics solution to mental causation face a trilemma depending upon whether they appeal to exact resemblance, rough resemblance, or resemblance in a certain respect. *If* they appeal to exact resemblance to construct classes of properties, then they must *either* concede that there are no mental properties (in which case, non-reductive physicalism is false) *or* that mental property instances are distinct from physical property instances (in which case, they have no solution to the problem of mental causation). That is, either they have no solution to the problem or the doctrine for which they sought to provide a solution cannot be formulated.

On the other hand, *if* they appeal to rough resemblance, or resemblance in a certain respect, they face one of two difficulties. If the appeal is to rough resemblance, then they face the problem of imperfect community and undermine the motivation for adoption of a trope metaphysics. That is not to presume that the problem of imperfect community cannot be resolved. Perhaps it can. Indeed, Resemblance Nominalists, or friends on their behalf, have suggested solutions which we do not have to (and don't have the space to) evaluate here (e.g., Hirsch 1993, 58–9; Rodriguez-Pereyra 2002, Ch. 9).

The point is that unless it turns out these solutions work only for tropes—which is unlikely—there is no particular reason for adopting a trope metaphysics rather than resemblance nominalism.

I guess it can be argued that, if a solution were available, then the appeal of resolving the problem of mental causation may make trope metaphysics independently attractive. A lot would turn on whether causation by events—say—rather than causation by property instances—preserves less of our intuitions concerning whether the mental is efficacious. I have argued elsewhere that proponents of the trope solution have to make versions of the same moves that they criticize in those who claim that mental causation just involves events and not properties (or their instances). They have to suggest that apparently intuitive claims about efficacy reveal something, instead, about the pragmatics of explanation (Robb 1997; Noordhof 1998, 225–6). In section 5, I will explain how the claims have a plausible ontological basis instead.

If trope metaphysicians appeal to resemblance in a certain respect—perhaps even exact resemblance in a certain respect—then they needn't face the problem of imperfect community but the respects (be they mental or physical) allow the problem of mental causation to be raised once more. It can be legitimately asked, are mental property instances efficacious in virtue of their mental respect or their physical respect? The trope metaphysician cannot deny this structure because they have appealed to it to resolve the problem of imperfect community (Gibb 2004, 473–5).

We must conclude that proponents of the trope solution fail to establish that a successful response to Kim's argument constitutes a successful defence of the claim that mental properties are causally relevant. Nor does their proposal sidestep the problem I raised with regard to the application of the subset approach. It indicates that we need to recognize the existence of mental tropes in addition to narrowly physical tropes (and their arrangement) to capture the additional causal powers that mental properties possess.

In this section, I have examined how a certain ontological response to Kim's argument—identity of property instances—leads to distortion of what we should say about the causal powers of mental property instances, implausible theses concerning instance identity, or abandonment of the advantages of trope metaphysics. This is the second of the two over-reactions I identified. Before I turn to my own approach, let me briefly discuss an approach which self-consciously does not take BP seriously.

#### 4. Challenge from the Unilevellers

Unilevellers deny that the world is layered, the idea that there are different levels of properties. One unileveller position would be to adopt a trope metaphysics which takes properties to be fully determinate tropes in exact resemblance classes and concludes that, for example, while there are concepts of mental properties, and their instances, there are, in fact, no such properties or instances of them. Instead, our mental concepts capture families of *similar* tropes (see, e.g., Heil 2003, 140–3, 153; Heil’s modes differ from tropes in that the objects they characterize are essential to their identity).

If there are no mental properties, then they cannot be causally relevant. Nevertheless, unilevellers suppose that our mental discourse picks out something which is causally relevant. Take the case of pain experience of a particular type. The picture is captured by a negative and positive claim. First, the negative:

There is no single respect  $R$  (exact resemblance in some way) in virtue of which  $\{A_1(p_1, p_2, p_3 \dots), A_I(p_{I1}, p_{I2}, p_{I3} \dots) \dots\} \dots\}$  are all members of the class of property instances that fall under my concept of a particular type of pain experience (hereafter, pain experience instances) (cf. Heil 2003, 153).

Second, the positive claim:

‘A particular instance of pain experience caused a particular instance of writhing’ is true in virtue of there being some member of the set of pain experience instances, say a particular instance of C-fibre firing in way  $W$ , which caused some member of the set of writhing instances.

Our talk of pain experience of a certain type is causally relevant because, although there are no such properties, our concept of it has conditions of application which pick out a class of property instances one of which was efficacious in the circumstances.

I have two related objections to this position. The first is that it turns inferences about how creatures behave as a result of being in pain—or being in other mental states—into relatively shaky inferences. We are inclined to assert that if a pain experience of a certain type were instantiated in a silicon creature (a sentient robot), it would writhe. But exactly how it would behave is open to question if the chain of resemblances that bundle all pain experiences together allow for significant differences. Perhaps my pain

experience and a silicon creature's pain experience resemble each other in being distracting, but not in giving rise to certain bodily responses. Identical causal powers are not guaranteed across subjects.

Unilevellers may dissipate this initial worry by insisting that the required amount of similarity will include the robot version of writhing in circumstances in which c-circuit activity occurs. This is not guaranteed because, once you go approximate, there is always the possibility that there will be sufficient similarity without this element. However, let me for the sake of argument concede it.

That will still not deal with the particular case. Suppose I am currently having a pain experience. Then I can reasonably think, if my friend Robbie the Robot was experiencing this, he would be writhing about in agony. However, in thinking about this pain experience, I am thinking about (according to the unileveller picture) C-fibre firing in way W. This C-fibre firing would not cause Robbie to writhe about in agony. Unilevellers deny that there is any mental property apart from this upon which my belief may be grounded. So they are committed to holding that I have no grounds for the belief in question.

It might be argued that this upshot is intuitive. Humans do respond to pain in different ways. Different creatures are likely to do so even more. But these observations are compatible with shared causal profile (as I noted before). The causal profile of a property will, in different contexts, manifest itself in different ways. The unileveller position is more radical than this. According to them, there is no shared causal profile—at best, just an approximate similarity in causal profiles of different property instances.

This brings me to my second objection. As we shall see in the next section, what is required for the causal relevance of mental properties is not simply the efficacy of an instance of a mental property (nor for that matter the efficacy of an instance of something picked out by a mental concept) but a condition-relative general relationship between the instantiation of mental properties and their target effects. By denying that such properties exist, unilevellers give up on this requirement. As such, this is a point against them.

The unileveller position derives much of its motivation from an appeal to truthmaking. The basic idea is that if  $A_1(p_1, p_2, p_3 \dots)$  putatively metaphysically necessitates  $bp_1$ , then it counts as the relevant part of a truthmaker of sentences with terms putatively referring to  $bp_1$ . There is no need for  $bp_1$  to exist. There is, however, another dimension which  $A_1(p_1, p_2, p_3 \dots)$  seems

less well suited to provide. That is, the generality that we associate with  $bp_i$  expressed in the inferences we are inclined to make. The precise nature of those will occupy us in the next section of this paper. But, to summarize the concern in a phrase, unilevellers mistakenly emphasize truthmaking at the expense of inference-basing.

## 5. Property Causation

Kim's original challenge to non-reductive physicalism was that it made mental property instances inefficacious. An answer to his argument is a necessary condition for a defence of non-reductive physicalism against the charge of epiphenomenalism but it is not sufficient. In addition, we need an explanation of how mental property instances are efficacious in virtue of being *mental* property instances. The appeal to different patterns of variation seemed to draw on material that might be helpful in this regard but at the expense of neglecting the detail needed to defend the claim that mental property instances were efficacious. The instance identity strategy was more focused on the latter but the problems with the subset view started raising issues about the efficacy of the mental instances qua being mental, which were revealed in cross-subject judgements about the effects of a particular pain instance.

My proposal is an attempt to satisfy both requirements. It focuses both on what is required for a particular instance of a mental property, or indeed any broadly physical property to be efficacious, and also on the element of generality that shows that the instance is efficacious in virtue of being a mental property. It runs as follows.

F is a property cause of G if and only if

*Particularity:* part of the (minimal) necessitation-base for the instance of F causes part of the (minimal) necessitation-base for the instance of G.

*Generality:* (part of) each (minimal) necessitation-base of F is such that all its instantiations would cause (or in the case of indeterminism, raise the probability of) an instantiation of one of the (minimal) necessitation-bases of G if they were in some causal circumstances C—where C may vary for each kind of necessitation-base.

Let me comment on various elements of this proposal.

First, the appeal to necessitation-base is meant to be understood in terms of metaphysical necessity—just as with the characterization of non-reductive



physicalism. The insistence on a *minimal* necessitation-base is addressed to the following difficulty. If T is the necessitation-base for F, then so is T plus the whole world apart from T. But we wouldn't want to conclude that F is efficacious because of the efficacy of some feature of the world unrelated to T (or, indeed, some feature of T unrelated to the minimal necessitation-base for F if T is also not the minimal necessitation-base for F). I have characterized the minimal supervenience-base in a previous publication; talk of minimal necessitation-base just focuses on the key element (see Noordhof 1999a, 307).

The intuitive idea is straightforward. The minimal necessitation-base for F is all that needs to be instantiated for a particular way of instantiating F. It is meant to capture, in some sense, the thought that broadly physical properties are constituted from arrangements of narrowly physical ones, although these ways may vary. We know that a certain understanding of property constitution cannot be right. Complex universals (if they exist) cannot be composed from other universals. The classic example to illustrate this is Lewis' case of methane (Lewis 1986). It is not composed of four hydrogen universals and one carbon universal because there are not four hydrogen universals. We can't understand property constitution simply in terms of property instance constitution either. Instances of methane may be composed from one instance of the property of carbon and four instances of the property of hydrogen but even this does not work for variably realized properties. A single universal cannot be constituted in various ways, even if its instances can be. So variably realized universals can't be said to have other properties as constituents.

For properties, rather than property instances, it is better to appeal straight to the idea of minimal metaphysical necessitation—if it can be defended against counterexamples as I urge. When it holds, it seems to follow that there is nothing over and above arrangements of instances of properties needed to constitute an instance of the target broadly physical property. If there were something over and above arrangements of instances of these properties, then there would always be the possibility that the additional element *could* fail to occur, even if it *wouldn't* (given the psychophysical laws).

Of course, particular analyses of minimal necessitation-base—and the background idea of property instance constitution—may fail. But since the idea is natural and, more importantly, does not implicitly draw on

claims concerning the efficacy of the target broadly physical properties, we could safely take it as a primitive without concern that it vitiates the substance of the account of property causation.

Second, the appeal to ‘part of’ is to allow for the possibility that something may count as efficacious only in virtue of an element of it being efficacious. The fire burned because of the presence of air in virtue of the fact that oxygen is part of air. I shall discuss this no further here but it is relevant to the issue of the efficacy of mental properties for which externalism is true (see Segal and Sober 1991; Noordhof 1999a).

So much for a preliminary understanding of the first element. Let me now turn to the second element, that of generality: the every minimal necessitation-base clause. Of course, I am not the first to recognize the implicit generality. Anybody who has offered an account of causal relevance in terms of law has also done so (e.g., Fodor 1989; Segal and Sober 1991, 15). However, first, my proposal does not appeal to law because it is questionable whether there is a law if the pattern I have identified holds, and, second, those who offered such an account often failed to appeal to the idea of minimal necessitation too. Yet an appeal to the latter is also required.

Appeals to law by themselves struggle to explain whether correlation between broadly physical properties reveals that their instances are standing in a causal relationship. Broadly physical properties which are nomically, but not metaphysically, necessitated by arrangements of narrowly physical properties will have a true general statement concerning their co-occurrence even if the broadly physical properties are intuitively inefficacious (Segal and Sober 1991, 4–5). So something extra is needed. Either this can be part of the conditions under which the generality would count as a causal law, or it can be characterized independently. That these conditions are needed is not in dispute.

Turning to the first point, my condition bears most resemblance to an account which appeals to a *ceteris paribus* law relating F and G to capture the generality involved in causal relevance. A preliminary analysis of *ceteris paribus* laws is that there is a *ceteris paribus* law relating F and G, if and only if, for all R, where R realizes F, there are some conditions C, such that, whenever R & C, then G and it is nomologically possible that R without C (Fodor 1991, 23–4; Schiffer 1991, 6–7). If the second condition were not met, then the law would be strict. The possibility of R without C provides conditions in which the correlation between F and G fails.

An objection to this analysis is that *ceteris paribus* laws have what Fodor has dubbed ‘absolute exceptions’: realizations for F for which there are no circumstances C which, together with the realization, are sufficient for an instance of G. One way in which conditions may be unequal is if F is realized by a dud. Fodor accommodates this by allowing that F can figure in a *ceteris paribus* law if most of the time, it is not realized by duds for G, and for other properties, say H, with which it also stands in a *ceteris paribus* law, the dud realization does have circumstances in which it yields an instance of G (Fodor 1991, 27–8). Others respond to this objection by denying the existence of *ceteris paribus* laws (Schiffer 1991).

Whichever way one goes, the characterization of my generality condition does not, then, involve an appeal to laws. However, its motivation remains intact. If two properties are co-instantiated, then the effects of this instantiation may be due to one or the other of the properties. One famous illustration is the soprano’s singing of ‘my love’, at a certain pitch and loudness, causing the glass to crack. It is plausible that the soprano’s singing is an instance of that pitch, that loudness, and those words. Yet we would not conclude that the glass cracking occurred in virtue of those words. So how should we differentiate?

According to the generality condition, the property of involving the words ‘my love’ does not serve to explain the pattern of causal relations concerning glass crackings, taking into account different ways in which the property of involving the words ‘my love’ may be realized. If the generality condition holds for a certain property for a target effect, then we have such an explanation. The causal relevance of a property, and not just one of its instances, is hard to deny if, for every type of minimal necessitation-base of a property, there are circumstances in which an instance of that property always causes the target effect.

Consider the property of being rickety. If something is rickety, it is likely to collapse. However, whether or not something will collapse depends upon the precise circumstances in which it is located and the precise way in which the property of being rickety is realized. If for every way of being rickety, there are circumstances in which collapse follows, then we can say that a collapse followed in virtue of being rickety. Otherwise, talk of ricketiness, at best, figures in a *ceteris paribus* law. If it does not meet the generality condition, then, while various ways of being rickety may be causally relevant, ricketiness is not.

Closer to our interests to begin with, consider the case of pain. If pain is realized by C-fibres firing in me, together with, perhaps, the laws which govern their causal role, and these firing fibres are transplanted into our silicon friend, Robbie the Robert, we would expect no writhing to ensue. Now it might be urged that, in that case, what is shown to be efficacious is the way in which the pain is realized and not pain itself. Nevertheless, there is a fact unaddressed by this suggestion. It is that all of the realizations of pain (e.g., in the case of Robbie the Robert C-circuit activity) have conducive circumstances in which they give rise to writhing behaviour. This is not the case for other properties with regard to that kind of behaviour. So it is plausible that there is something about pain, and not just pain in such and such a type of creature, which is responsible for the link between pain and writhing. It is this which grounds the claim that pain is causally relevant for this behaviour.

Suppose that there is a lone species in which instances of pain do not cause writhing. Would that imply that human pain does not cause writhing?<sup>1</sup> It would not because, by limiting the question to human pain, the generality condition would only apply to necessitation-bases of human pain. Nor would it even imply that pain is not a property cause of writhing. That would depend on whether the way in which pain is realized in this lone species might also be realized in other creatures in which it did cause writhings, or whether there were conditions in the lone species in which writhing might be so caused.

The generality condition is also related to, but importantly distinct from, a distinction drawn recently between sensitive and insensitive causation. Often, the latter distinction is made within the context of taking causation to be difference making, something I discussed in the second section of this paper (e.g., Woodward 2006, 7). A causal relation is relatively insensitive—between particulars, or types of things—if the counterfactual dependence between the causal relata holds in a variety of different background conditions. It is sensitive if this dependence is easily disrupted. Christian List and Peter Menzies extend this idea to include sensitivity, or otherwise, to the way in which the properties standing in the putative causal relation are realized. Cases of sensitive causation in this sense are taken to be counter-examples to the exclusion principle I mentioned earlier. List and Menzies

<sup>1</sup> This question was asked by an anonymous referee.

hold that both the way in which a property is realized, and the instance of the property, are to be counted as causes in such cases (List and Menzies 2009, 491–2, 497–9; in Menzies 2008, he seems committed to an exclusion principle to which this kind of case is a counterexample).

To illustrate, suppose that a certain kind of pain,  $P_a$ , has four necessitation-bases  $N_1$ ,  $N_2$ ,  $N_3$ , and  $N_4$  and let  $B_g$  be the utterance ‘That hurts!’ Suppose that, further, the following counterfactuals were true.

If  $P_a$  were not instantiated in  $S$ , then  $S$  would not utter  $B_g$ .

If  $N_1$  were not instantiated in  $S$ , then  $S$  would not utter  $B_g$ .

For the latter to be true, the closest worlds in which  $N_1$  is not present are ones in which  $B_g$  wouldn’t occur even though there is a replacement,  $N_2$ , and  $P_a$  is, thus, present. In those circumstances, List and Menzies conclude that both  $P_a$  and  $N_1$  are causes of  $B_g$ .

I can see why it is plausible to suppose that  $N_1$  is a property cause in that situation. It is far less clear why it is plausible to suppose that  $P_a$  is. Given what has been previously argued, we are allowed the question: Does  $N_1$  cause  $B_g$  partly in virtue of necessitating  $P_a$ ? Evidence that it is not in virtue of  $P_a$  is that, when a substitute realization,  $N_2$ , is present,  $B_g$  does not occur. List and Menzies suggest that the relationship between  $P_a$  and  $B$  is sensitive, depending upon the precise way in which  $P_a$  is realized. Instead, the sensitivity supplies evidence that it is  $N_1$  rather than  $P_a$  that is the causally relevant property. If the sensitivity were just the result of a failure of the right causal circumstances, then the case List and Menzies cite would not be a problem. The verdicts of the two approaches would coincide. The difference stems from the decision to count as one source of sensitivity the way in which  $P_a$  is realized. It is here that I think their account yields counterintuitive verdicts. Sensitivity is not compatible with causal relevance.

A consequence of my favoured account is that it delivers the verdict that there are causal relations between broadly physical properties. Kim’s argument may be viewed as questioning this on the grounds that all the work is being done by the arrangements of narrowly physical properties. Since my proposal does not make instances of broadly physical properties identical to instances of arrangements of narrowly physical properties, I don’t have an immediate response to this worry. True my proposal may get the right verdict in the sense of what we want to believe but the charge is that it shouldn’t.

Part of my response to this objection is contained in my reply to the challenge of the unilevellers. At this point, I emphasize another issue. The debate in this area begins by conceding that there are broadly physical properties but then challenges their existence by arguing that they are inefficacious. However, the initial concession undermines the challenge. Either you don't think broadly physical properties exist, in which case I draw your attention to the way in which they back inferences we want to make about how things will behave in different circumstances. Or you accept that they do, barring an argument to the contrary. Appeal to causal considerations will not provide such an argument because, in allowing that broadly physical properties exist, you must also allow that broadly physical causal relations exist. They are just one more species of property whose existence we have allowed as a result of their necessitation by arrangements of, in this case, causal relations between narrowly physical properties. There seems an entirely unmotivated asymmetry in the debate whereby causal relations are treated differently to any other kind of property. One illustration of this last point is that, just as other properties seem to stand in relations of determinable to determinate, so do various types of causal relations, for example, 6 inch diameter ball depression, ball depression, depression, specify causal relations at different degrees of generality. These are determinable causal relations in which determinable properties may stand.

There might be other reasons to resist the claim that broadly physical causal relations exist. My point is simply that these considerations had better not take the same form as considerations, independent of causation, for rejecting the existence of broadly physical properties in general. We were supposed to be provided with a consideration from causation against the latter, not just a blanket favouring of the narrowly physical. My account of property causation is an attempt to identify when these broadly physical property causal relations are present, and how they capture something in addition to particular arrangements of narrowly physical properties, through the generality condition.<sup>2</sup>

Another objection to the proposal discussed recently derives from the possible truth of a powers ontology. A powers ontology takes the causal profile of a property to be internal to it. By that I mean that the causal profile of the property does not depend upon laws which hold, in addition, but

<sup>2</sup> This paragraph was written in response to an objection by an anonymous reviewer.

rather given that the property is instantiated, certain laws hold. Suppose that emergent dualism is true. Then one element of the causal profile of an arrangement of narrowly physical properties is that they cause the presence of an emergent non-physical property. If arrangements of narrowly physical properties in such an ontology could not fail to have their causal profiles, then it follows of metaphysical necessity that, if the arrangement of narrowly physical properties is present, then the emergent dualist property is instantiated. Nevertheless, it could still be the case that it is not part of the causal profile of the emergent dualist property that it cause some target effect which is part of the causal profile of the arrangement of narrowly physical properties. Indeed, that is what epiphenomenal emergent dualists assert. The objection to my proposal is that it suggests a certain account of how broadly physical properties can inherit the efficacy of arrangements of narrowly physical properties that cannot allow for this possibility (O'Connor 1994, 97; Wilson 2005, 436–47).

One response to the specifics of the objection is to say that if a powers ontology were true, there would be no basis for being an epiphenomenal emergent dualist. The grounds for being a dualist are usually the intrinsic features of phenomenal states. If a powers ontology were true, there would either be no intrinsic features, or the intrinsic features in question would not be different for narrowly physical properties. I mention this last possibility to take into account C. B. Martin's position that every property has both a qualitative and dispositional aspect (e.g., Martin 1997).

Nevertheless, this does not deal with the general structure of the objection. Suppose that there is a property  $C_1$  which has a causal role  $CR_1$  which includes, if  $C_1$  is instantiated in  $S$ , then  $E_1$  and  $F_1$  is instantiated. Then  $C_1$  and  $S$  metaphysically necessitate  $E_1$  and they also metaphysically necessitate  $F_1$ . Doesn't my position have as the upshot that  $E_1$ 's causal role  $ER_1$  must include the instantiation of  $F_1$  when, intuitively, it need have nothing to do with the instantiation of  $F_1$ ?

Here are two more general lines of response. First, my talk of metaphysical necessitation was meant to capture the important characteristic of previous talk of constitution, namely that if that which was necessitated by the necessitation-base involved nothing more than what was in the base, then given the base, the necessitated must also be instantiated. It might be argued that, if a powers ontology is true, metaphysical necessitation cannot suffice to capture our notion of constitution even given the assumption that

the entities it associates are contingent. In which case, we might take property instance constitution as a primitive and note that it supports metaphysical necessitation claims but is not the only possible support. The proposal would be reformulated in terms of constitution. This is not a particularly damaging adjustment because there is no reason to think that the proper understanding of property instance constitution must appeal to causation or kindred notions that I am seeking to illuminate by my proposal.

Second, we can deny that a powers ontology implies dispositional essentialism, the view that the causal profile of a property is essential to it. In which case, there is no reason to accept that  $C_1$  and  $S$  metaphysically necessitate  $E_1$  and hence no grounds for supposing that my proposal must accept the verdict that  $E_1$  has the instantiation of  $F_1$  as part of its causal profile. The point is especially plausible with regard to the fundamental laws relating arrangements of narrowly physical properties and the properties put forward by epiphenomenal emergent dualists as mental properties. They envisage that these stand in isolation from other narrowly physical properties and so it is perfectly conceivable that the same narrowly physical properties may be instantiated without this part of their causal profile. However, more generally, any particular aspect of the causal profile of a property could be plausibly supposed to be absent with the rest still present.

One way to think of these possibilities is in terms of counterpart theory. We can suppose that, in other possible worlds, there are properties with a strong similarity to the causal profiles instantiated in our world—structurally speaking—and yet some differences. The question arises whether it is plausible to consider these properties counterparts of the property in our world. It is hard to see why not. Properties with different causal profiles may be counterparts and yet nothing that has been said rules out the possibility that the causal profile is internal to the property. Indeed, counterpart theory was introduced to, amongst other things, deal with the problem of accidental intrinsic properties of particulars. The suggested strategy just applies this to the case of properties (for further discussion, see Noordhof 2010).

A second objection to the proposal follows from something I said earlier. I remarked that the causal powers of many broadly physical properties exceeded those of the arrangement of narrowly physical properties that metaphysically necessitated them. This can seem wrong on one of two counts. First, how can it be that novel causal powers are metaphysically necessitated by arrangements of properties which, it is alleged, individually



or together don't possess them? Second, if it is allowed that they do metaphysically necessitate novel causal powers for the sake of argument, then why doesn't just admitting this make the causal powers accrue to the arrangements of narrowly physical properties which do the necessitating?

The answer to the second question is that the powers of  $BP_1$  don't transmit to  $A_1(p_1, p_2, p_3 \dots)$  because, first, the instance of  $BP_1$  is not identical to, nor caused by, the instance of  $A_1(p_1, p_2, p_3 \dots)$  and, second, downward transmission of causal powers does not apply because  $A_1(p_1, p_2, p_3 \dots)$ ; the latter stops some of the powers of  $BP_1$  from being manifestable, namely those associated with other physical realizations of  $BP_1$ . Of course, part of the causal profile of a property  $F$  need not be manifested in order for the property to have that causal profile. However,  $F$  cannot have, as part of its causal profile, the potential for causal relations it could not stand in while remaining the property it is, given the laws which hold. I mention this second point in case it is thought that a version of my proposal should explain how efficacy of broadly physical properties transmits downwards even if the attribution of the causal powers is not immediate in virtue of the first point.

This response to the second question makes it harder to see how one could provide an answer to the first. How can a particular arrangement of narrowly physical properties necessitate a property which has causal powers more extensive than it? An incomplete answer would be that, although the powers of  $BP_1$  exceed particular minimal necessitation-bases of it, if we consider all the various minimal necessitation-bases, then the complete set of causal powers that these minimal necessitation-bases have is possessed by  $BP_1$ . There are two problems with this response. The first is that its plausibility partially rests upon the assumption that all the possible minimal necessitation-bases of  $BP_1$ , which give it distinct causal powers, are instantiated in a particular world. In the absence of this, upon what basis could we conclude that the other elements of the causal power were present? This is not merely a notional objection. Many candidate Bps actually allow for physical and non-physical minimal necessitation-bases with the presumption that there are none of the latter if physicalism is true. Second, even if we have some explanation of why we may allow that *all* the powers associated with  $BP_1$  are instantiated, it is unclear why we should conclude that they are instantiated with regard to a particular instance of  $BP_1$  when necessitated by a minimal necessitation-base that cannot have at least some of the powers.

The proper response is to distinguish *constitution as co-ordination of the small* in making up the bigger from *constitution as involving grounding*, in which the constituents are viewed as fundamental. These are clearly distinct notions otherwise we would have a fast argument from something being a constituent to monism (the priority of the whole) being false. We should reject the idea that arrangements of narrowly physical properties *constitute* broadly physical properties and, more specifically, that the causal relations of the narrowly physical properties so arranged *constitute* the causal relations of the broadly physical properties in a metaphysically fundamental sense in which the constituents are taken to be primary. Instead, the proper relationship between narrowly physical and broadly physical properties is one of harmonization (see Noordhof 2003, 105–6). The right metaphor is not of an economical God who, if only he were to fix the arrangements of the physical, he would have the broadly physical properties fixed, but rather of a God subject to constraints. He is not allowed to instantiate some of the first lot without instantiating some of the second lot too. Broadly physical properties, and their causal relations, are no less fundamental than the arrangements of narrowly physical properties with which they are closely related.

From this alternative perspective, the relations of metaphysical necessitation between the arrangements of narrowly physical properties and broadly physical properties capture the constraints upon instantiation, and co-instantiation, between these properties. If the constraints are not observed, then the causal relations of the properties would literally be incompatible with each other. We would have an impossible world. Arrangements of narrowly physical properties only appear to be ontologically fundamental because the causal relations identified at that level are more detailed than those identified between broadly physical properties. Since there are various ways more general causal relations may be realized by more detailed causal relations, we have an asymmetry. Arrangements of narrowly physical properties fix what broadly physical properties there are but the latter only imply that one or other of various arrangements of narrowly physical properties are present. However, interpreting this asymmetry as implying that arrangements of narrowly physical properties are fundamental is not mandatory if constitution is just co-ordination of the small.

Of course it is true that instances of narrowly physical properties may be present, and stand in causal relations, whether or not broadly physical

properties are present. That might suggest that they have some priority. However, once the alternative picture is in play, this fact needs to be set in the context of other observations. First, since broadly physical properties may be related to different arrangements of narrowly physical properties and, indeed, in some cases, to arrangements of non-physical properties, there is no reason to take broadly physical properties to be dependent on their instances' actual constituents. Second, arrangements of narrowly physical properties are subject to constraints on co-instantiation stemming from their constituting, in the co-ordination of the small sense, broadly physical properties. If a broadly physical property is to be instantiated with certain causal powers, and certain constituent instantiations of narrowly physical properties are to be instantiated, then certain other constituent instantiations of narrowly physical properties must be instantiated too, namely those implied by the instantiation of the broadly physical property with those other narrowly physical properties as constituents.

Recognition of novel causal powers, in the way that I have sketched, does not constitute a rejection of a weak causal closure principle like 'every event with a cause has a narrowly physical cause' for, at least, two reasons. First, that principle is compatible with there being non-physical causes too. But, second, and more important in the present context, denying that arrangements of narrowly physical properties constitute the causal relations of broadly physical properties, does not mean that there are events with broadly physical properties as causes without arrangements of narrowly physical properties as causes. Allowing that there are cases in which the broadly physical properties are no less fundamental than narrowly physical ones does not imply that they have causal consequences without arrangements of narrowly physical properties being present.<sup>3</sup>

## 6. Concluding Remarks

The causal relevance of properties, or property causation as opposed to property instance causation, turns on two issues: first, causal facts about their instances; second, the causal significance of a generality captured in terms of the properties in question. Focus on difference making, or patterns of

<sup>3</sup> This paragraph was written in response to a question by an anonymous referee.

variation, are better at capturing the second element but they ignore the first element at their peril. Even with regard to the required display of generality, they fail to observe the conditions which should be met for a property to be said to be causally relevant (rather than just its instances). My own proposal—involving a particularity condition plus a generality condition—has the merit of indicating how the two elements should be integrated. It suggests that the concern about efficacy, within the context of non-reductive physicalism, partly stems from an inadequate understanding of how this integration should be undertaken given that there are different minimal necessitation-bases for broadly physical properties together with an unmotivated asymmetry in the treatment of causation itself, as just one kind of property amongst others. The other root of the trouble is the focus on developing a picture of the world which places the emphasis on truthmaking rather than inference-basing. The recognition that non-reductive physicalists should allow broadly physical properties to have causal powers which outstrip their bases in a circumscribed sense and, as a result, resist the constitution-as-grounding assumption, suggests a different understanding of how we should see the relationship between broadly and narrowly physical properties. If this proposal is along the right lines, it provides support for the view that the apparent problem of mental causation is only properly resolved if the metaphysical picture, in which it might figure, is made much clearer: the theme of the AHRC funded project to which this paper was a contribution.<sup>4</sup>

## References

- Campbell, K. (1981). 'The Metaphysic of Abstract Particulars'. *Midwest Studies in Philosophy*, 6: 477–86.
- (1990). *Abstract Particulars*. Oxford: Basil Blackwell.
- Ehring, D. (1997). *Causation and Persistence*. Oxford: Oxford University Press.
- Fodor, J. A. (1989). 'Making Mind Matter More'. *Philosophical Topics*, 17(1): 59–79.
- Repr. in Fodor (1990), *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press: 137–59.

<sup>4</sup> Many thanks to the audience of the Metaphysics of Mind conference in Durham to which an earlier version of this was presented, Sophie Gibb, for inviting me to participate and, more recently, to the members of the York Mind and Reason group who subjected my ideas to scrutiny: Keith Allen, Will McNeill, Ema Sullivan-Bissett, and Rachael Wiseman.

- (1991). ‘You Can Fool Some of the People All of the Time, Everything Else Being Equal; Hedged Laws and Psychological Explanation’. *Mind*, 100(1): 19–34.
- Gibb, S. (2004). ‘The Problem of Mental Causation and the Nature of Properties’. *Australasian Journal of Philosophy*, 83(3): 464–76.
- Gibbons, J. (2006). ‘Mental Causation without Downward Causation’. *Philosophical Review*, 115(1): 79–103.
- Heil, J. (2003). *From an Ontological Point of View*. Oxford: Oxford University Press.
- Hirsch, E. (1993). *Dividing Reality*. Oxford: Oxford University Press.
- Kim, J. (1992). ‘Multiple Realization and the Metaphysics of Reduction’. *Philosophy and Phenomenological Research*, 52: 1–26. Repr. in Kim (1993), *Supervenience and Mind*. Cambridge: Cambridge University Press: 309–35 [page references in text to latter].
- (1998). *Mind in a Physical World*. Cambridge, MA: MIT Press.
- (1999). ‘Supervenience Properties and Micro-Based Properties: A Reply to Noordhof’. *Proceedings of the Aristotelian Society*, 99: 115–18.
- (2005). *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.
- Lewis, D. (1986). ‘Against Structural Universals’. *Australasian Journal of Philosophy*, 64: 25–46.
- List, C. and P. Menzies (2009). ‘Nonreductive Physicalism and the Limits of the Exclusion Principle’. *Journal of Philosophy*, 106(9): 475–502.
- Manley, D. (2002). ‘Properties and Resemblance Classes’. *Noûs*, 36(1): 75–96.
- Martin, C. B. (1997). ‘On the Need for Properties: The Road to Pythagoreanism and Back’. *Synthese*, 112(2): 193–231.
- Menzies, P. (2008). ‘The Exclusion Problem, the Determination Relation, and Contrastive Causation’. In J. Hohwy and J. Kallestrup (eds.), *Being Reduced: New Essays on Reductive Explanation and Special Science Causation*. Oxford: Oxford University Press: 196–217.
- Noordhof, P. (1998). ‘Do Tropes Resolve the Problem of Mental Causation’. *Philosophical Quarterly*, 48 (191): 221–6.
- (1999a). ‘Causation by Content?’. *Mind and Language*, 14(3): 291–320.
- (1999b). ‘Micro-Based Properties and the Supervenience Argument: A Response to Kim’. *Proceedings of the Aristotelian Society*, 99: 109–14.
- (1999c). ‘The Overdetermination Argument versus the Cause-and-Essence Principle—No Contest’. *Mind*, 108(430): 367–75.
- (2003). ‘Not old . . . but not that new either: Explicability, Emergence and the Characterisation of Materialism’. In S. Walter and H.-D. Heckman (eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. Charlottesville: Imprint Academic: 85–108.

- (2010). ‘Emergent Causation and Property Causation’. In C. Macdonald and G. Macdonald (eds.), *Emergence in Mind*. Oxford: Oxford University Press: 69–99.
- O’Connor, T. (1994). ‘Emergent Properties’. *American Philosophical Quarterly*, 31(2): 91–104.
- Olson, E. (1997). *The Human Animal*. Oxford: Oxford University Press.
- (2007). *What Are We?* Oxford: Oxford University Press.
- Robb, D. (1997). ‘The Properties of Mental Causation’. *Philosophical Quarterly*, 47: 178–94.
- Rodriguez-Pereyra, G. (2002). *Resemblance Nominalism*. Oxford: Oxford University Press.
- Schaffer, J. (2005). ‘Contrastive Causation’. *Philosophical Review*, 114(3): 327–58.
- Schiffer, S. (1991). ‘Ceterus Paribus Laws’. *Mind*, 100(1): 1–17.
- Segal, G. and E. Sober (1991). ‘The Causal Efficacy of Content’. *Philosophical Studies*, 63: 1–30.
- Shoemaker, S. (1981). ‘Some Varieties of Functionalism’. *Philosophical Topics*, 12(1): 83–118. Repr. in Shoemaker (1984, 2003), *Identity, Cause and Mind*. Oxford: Oxford University Press: 261–86.
- (2007). *Physical Realization*. Oxford: Oxford University Press.
- Steward, H. (1997). ‘On the Notion of a Cause “Philosophically Speaking”’. *Proceedings of the Aristotelian Society*, 97: 125–40.
- Whittle, A. (2007). ‘The Co-Instantiation Thesis’. *Australasian Journal of Philosophy*, 85(1): 61–79.
- Wilson, J. (2005). ‘Supervenience-based Formulations of Physicalism’. *Noûs*, 39(3): 426–59.
- Woodward, J. (2006). ‘Sensitive and Insensitive Causation’. *Philosophical Review*, 115(1): 1–50.
- Yablo, S. (1992). ‘Mental Causation’. *Philosophical Review*, 101(2): 245–80.