In Geoffrey Brennan, Robert Goodin, and Michael Smith (eds.) *Common Minds: Essays in Honour of Philip Pettit*. Oxford: Oxford University Press, 2007: 1-27. References should cite the published version.

Beyond Program Explanation[1]

In a number of articles over the years, and especially in *The Common Mind*, Philip Pettit has defended a distinctive and influential solution to the so-called 'problem of mental causation' – the problem of how minds can make a causal difference to the world. This view has become known as "program explanation", hereafter PE. As its name indicates, PE is more than just a proposal purporting to solve the problem of how our minds can influence events in the physical world. It is also an account of how the 'special' sciences can be explanatorily autonomous. The 'problem' addressed by PE is multifaceted, and we cannot deal with all of its aspects here. In what follows, we want to concentrate on that part of his proposal that deals with issues concerning mental causation, since *inter alia* PE is meant to demonstrate how mental properties can be causally relevant to the behaviour they explain. We begin by outlining the problem of mental causation that concerns us. We then give a brief account of Pettit's solution, indicating what we take to be its principal shortcomings, before providing what we take to be the correct solution. We develop our criticism of PE by elaborating on the metaphysics of the preferred alternative. Finally, we reinforce our view by showing how it can be deployed to defeat a powerful argument, mounted by Jaegwon Kim, against the possibility of causally relevant emergent properties.

1. Introducing The Problem

What is the problem that appeal to PE is meant to solve? Classically, the mental causation problem arises as a consequence of a commitment to non-reductive

[1] Versions of parts of this paper have been read at a number of venues, including the Research School of Social Sciences at The Australian National University, the Australasian Philosophical Association (NZ Division) Conference at the University of Otago (1999), The South African Philosophical Association Conference (2000), and the University of Massachusetts (Amherst). We are grateful to audiences for their comments. We are also indebted to Tim Elder and Michael Smith for their comments on the penultimate draft.

physicalism, together with some supplementary assumptions concerning causation.[2] The thesis of physicalism entails a monist ontology, which we take minimally to involve the claim that all events are physical events.[3] The non-reducibility claim requires acknowledging that some of these events are also mental events, which we take to be the claim that although every event that has a mental property has a physical property, the mental and physical properties are irreducibly distinct. There are two relevant assumptions concerning causality in play. The first is that mental events, specifically, believings, desirings, hopings, intendings, can cause physical effects. The second is that physical effects have sufficient physical causes. The question that immediately arises is: does the 'mentality' of the mental events contribute to the production of effects thus caused? Prima facie it would appear not; the 'physicality' of those physical causes does all the work.

This problem is sharpened if one accepts, as Pettit seems to do (and we certainly do), a certain metaphysical view of the nature of events involved in the causal transaction. This metaphysics requires that events be exemplifyings or instancings of properties in objects at, or during intervals of times. This is the Property-Exemplification Account of events (hereafter PEA), one that we describe in more detail later. According to it, a mental event just is the exemplifying of a mental property in an object at a time, and also has various properties, such as the property of being an exemplifying of a mental property in an object at a time. For example, the event which is my thinking of Vienna now just is the exemplifying in me now of the property, *thinks of Vienna* (a property of me), and that event has the property of being a thinking of Vienna. Similarly, a physical event just is the exemplifying of a physical property in an object at a time, and it too will have various properties. Given the physicalist commitment, each mental event, i.e., exemplifying of a mental property, will be (i.e., be identical with) an exemplifying of a physical property.[4]

---

[2] 'Non-reductive physicalism' is anomalous monism without the commitment to there being no psychological or psychophysical laws. We avoid this commitment just because the notion of 'law' at stake is controversial. Nothing hangs on this issue in this paper.

[3] We speak here of events, but the thesis is intended to cover all mental phenomena, however they are to be analyzed in terms of events.

[4] In the terminology preferred by Kim, whose version of the account we describe here and develop further in section 4, events are *exemplifications* of properties by objects at times. But Kim himself, and many others who take a universalist rather than a tropist view of properties, often use the terms 'instance' as alternatives to the term 'exemplification' (and thus claim, for example, that a mental event is an instance of a property at a time in an object). We ourselves prefer 'exemplifyings' to 'exemplifications' (along the lines of Lombard (1986)), since it makes clear that events are fundamentally changes, whose 'constitutive' properties are dynamic rather than static, or its cognate

Further, events themselves have properties, and have them by instantiating them. Now the problem looks like this: can the mental property of a mental/physical event exercise any causal influence? The assumption that physical effects have physical causes now has the consequence that, in any case where there is putative mental causation of a physical effect, the physical *property* of the mental/physical event must be the one in virtue of which that event causes the physical effect. The mental property looks inert; the only way of rescuing any causal influence on its behalf, so it seems, is to make the physical effect over-determined – brought about both in virtue of the mental/physical event's being an instance of the physical property and in virtue of that event's being an instance of the mental property. Of course there may well be some cases of causal over-determination, but such *systematic* over-determination is exceedingly implausible, so this way out looks desperate. Another way is to secure the causal effectiveness of the mental property of the event by identifying it with a physical property. But this is a reductive move that looks increasingly implausible; and, in any case, requires the rejection of the position we wish to defend, non-reductive physicalism. Epiphenomenalism looms.

This is the problem that appeal to PE is meant to solve. Or so one might suppose, given Pettit's concentration on the task of making what he calls 'higher-order' intentional states causally relevant to the production of behaviour. In fact, though, Pettit sees the problem as two-fold, and we think that part of his reluctance to accept an alternative, metaphysically less baroque, solution lies in his conflation of

---

term, 'instancings'; since we think that failure to do so blurs the crucial distinction between a substance and an event. Given the universalist (as contrasted with a tropist) view of properties, according to which an exemplification/instance of a property just is the thing that has it, we would have to say that *Jones* is the instance of the property, *runs*, since, according to the property-exemplification account, as developed by Kim, this is a property of Jones, and so is a constitutive property of the event which is Jones's running. But although Kim wants to say that the subject of that event is Jones, the exemplification of the property, *runs* by Jones is an *event*, a running, not the event's subject. We can avoid this problem altogether if we distinguish instances from instancings (i.e., exemplifyings), since we can then maintain (1) that an instance of a property is the thing that has it (whether this is an object or an event), (2) that events just are (i.e., are identical with) exemplifyings of dynamic properties of objects in those objects, *and* (3) that an instance of a property of an event just is the event that has that property. Events, like any other entities, have properties by instantiating them, but their constitutive properties are not, according to PEA, properties that they possess. These distinctions are important to our solution to the problem of causal relevance, since only certain ways of developing the PEA will make that solution possible. We return to this in section 4. For more on the distinction between static and dynamic properties, and the differences between Kim's and Lombard's versions of the PEA, see Macdonald (1989).

We would prefer to avoid the term 'instances' entirely, since it suggests a trope view of properties, which we reject. But, since many parties to the dispute concerning the problem of mental causation, including Pettit and Kim (esp. Kim 2003), regularly talk of events as instances of properties

the two features of the problem, hinted at above. The problem, which he introduces in the form of the question, "What is the relation between a higher-order [believing that *p*] and lower-order cause in virtue of which the higher-order counts as causally relevant?" (Pettit 1993: 33)), is divided into two parts. First, there is "the state played a certain causal part" and, and second, there is "it did so in virtue of the property of being a belief that *p*" (Pettit 1993: 34).[5] We take the first feature (how the state can play a 'causal part') to be the problem of causal *efficacy*, and the second to be the problem of causal *relevance*.[6] Efficacy has to do with causes and effects extensionally conceived. Given that *c* caused *e*, any descriptions referring to *c* and *e* will (suitably arranged) yield a true causal statement. This means that a description of *c* can pick out *c* by specifying *any* property *c* possesses, and likewise for *e*. The extensionality of the causal relation ensures that *whatever these properties are*, the resulting causal statement will be true. Given the profusion of properties possessed by all events (think of 'mere' Cambridge properties), it is clear that not all true causal claims will yield causally illuminating *explanations* of why *e* occurred. This is why there is an issue of causal relevance, as distinct from causal efficacy: only some

---

– intending the *universalist* view of properties as multiply-exemplifiable entities that can be (wholly) present in many places at the same time – we will, for present purposes, speak in these terms too.

[5] Again, Pettit and others speak of mental properties as higher-order properties, where we would use the term 'higher-level'. Higher-level properties should not be confused with higher-order ones. Higher-order properties are properties of properties, not properties of the things that have them in virtue of their possession of other properties. It is common, especially in functionalist treatments in the philosophy of mind, to use 'higher-order' rather than 'higher-level' when talking about mental properties such as being in pain, or dispositional properties like solubility. It's important that it's quite different from the contemporary logician's usage (though similar to Russell's and Ramsey's). In contemporary terms, 'solubility', like 'being a number', is a first-order predicate and so stands for a first-order property because its instances are particulars. However, both predicates might be classed as impredicative, i.e., specificiable by phrases that include second-order quantification over all properties, including those properties themselves. Thus, 'λx(*x* is soluble)' might be specified by something like, '$\exists F(Fx \,\&\, \forall y(Fy \,\&$*y* is placed in water$\rightarrow$*y* dissolves)', where we have a second-order quantifier, '$\exists F$', which ranges over all properties, including solubility (just as the bound variable in '$\iota x(\forall y(x \mid y \rightarrow x$ is taller than *y*)' impredicatively specifies the tallest person. Pettit and others who are functionalists with regard to mental properties use the term 'higher-order' precisely because, in order to give the definitions of such properties, we need to use higher-order quantification. They take such properties as being in pain, for example, as higher-order because it is the state of being in a state that is causally related to others in the way that is characteristic of pain. However, for reasons just given, we take mental properties of events to be higher-level properties – properties of events that have them in virtue of their possession of other properties, rather than properties of properties of events. Be that as it may, since many of the examples used by Pettit and others in the mental causation debate concern both higher-level and higher-order properties, and since the issue of realization of one property by another that is central to the debate seems to arise with regard both sorts of properties, we will, for present purposes, use the term 'higher-order' to cover both sorts of case.

[6] Things get a bit confusing, however, since, although Pettit speaks of properties and their instances, as we do, sometimes his talk of states is elliptical for talk of property-instances, and sometimes his talk of states is elliptical for talk of properties. Here it looks like he means, by 'state', 'event' (i.e., 'property-

properties of the event that is the cause are causally relevant to the production of the effect, so only *those* properties, suitably specified, will yield causal explanations of the effect. Keeping this distinction in mind we can ask: which aspect of the mental causation problem does the PE strategy solve? The problem of causal efficacy, the problem of causal relevance, or both? We think that Pettit is more concerned with causal relevance, but this leads him to adopt an unusual, and ultimately very implausible, approach to the problem of causal efficacy. Or so we shall argue.

2. The Program Explanation Solution

Fundamentally, the PE strategy construes the notion of a mental property's "determining an effect" as non-causal. PE thus bites the bullet: mental (and presumably all other special science properties) are *not* causally efficacious, in the sense that events that are instances of them do *not* bring about the effects they do in virtue of being instances of such properties. Mental properties are taken to be higher-order properties that supervene on physical properties of events. In any case where a mental property is thought to be causally efficacious in the production of an action (in the sense just specified), what really happens is that the instantiation of the higher-order (mental) property "ensures that" a lower-order (physical) property is instantiated, this lower-order-property doing the causal work (again, in the sense that the event that is an instance of that lower-order property brings about the action in virtue of being an instance of that property). As Pettit puts the point,

> The general idea in the program model…is that a higher-order property is causally relevant to something when its instantiation ensures or at least probabilifies, in a non-causal way, that there are lower-order properties present which produce it. (Pettit 1993: 37)

So an instantiation of a mental property will "program for" the instantiation of those physical properties required for the production of the physical effect. The "ensuring that" and "programming for" are non-causal relations so there is no causal competition between mental and physical properties, and so no over-determination.

In addition to solving the over-determination problem, the PE model is said to have the virtue of presenting intentional causation as like many other cases of higher-

---

instance'), but, as we shall see, at other times he seems to mean 'property'. See especially our

order causation. To this end the PE model is supported by various examples in which it is alleged that one finds just such a higher-order programming for lower-order properties whose instantiations then cause the effects that one may have thought were caused by (instantiations of) the higher-order property. Thus, consider the eraser's elasticity, which enables it to bend. Its elasticity is a higher-order property, realized by a lower-order property, its having a certain molecular structure. The elasticity is said to non-causally program for its realization by a relevant molecular-structural property, instances of this lower-order property then producing the bending effect. Thus, it is claimed, "The dispositional state programs for the bending." (Pettit 1993: 39)[7] We return to this example later.

The PE strategy and model works with a number of assumptions, four of which are salient to our discussion. First, the model assumes that the lower-order property that *is* causally effective need not be the most immediate lower-order property; it may be one further down the ladder, so to speak. That is, the model is agnostic concerning the level at which causal efficacy resides. Second, the strategy does not take on board the problem of showing how causation works; it is intended to be compatible with most theories of causation. Third, the model is to be interpreted as an ontological one. "It is an account of how, in the objective arrangement of things, higher-order causes relate to their lower-order counterparts. It is not an account of how we must subjectively come to know that certain higher-order causes are causally relevant." (Pettit 1993: 35) Fourth, the strategy takes as its metaphysical basis the distinction between properties and their instances: properties are abstract and universal, their instances are worldly, and so only instances of properties are involved in causal transactions.

Brief comments about these background assumptions are in order. We accept the first – the causal 'grounding' of a higher-order property need not take place at an order immediately 'below' it. However, the second is problematic. On some accounts of causation, the strategy does not work. This becomes apparent when we cite our main complaint against the PE model, in section 3. We accept the ontological constraint of the third assumption, and also the specific metaphysical commitments made by the fourth assumption. Any account of mental causation should be as

---

discussion of his examples in section 3.

ontologically robust as are accounts of any other kind of causation. Otherwise, it risks not being about what it sets out to be an account of, the obvious fact that our minds are causally active (by which we mean here, so as not to beg questions, that mental properties are causally relevant to the events that instances of them 'produce'). The metaphysics of properties as universals has been challenged, but it is common to both PE and our preferred solution, so it will not be questioned here.

3. The Problem with Program Explanation

With causal explanation we assume that the cause cited in the explanation is crucial to any account of how it works as an explanation. Similarly, the PE strategy will work only if the 'program' part of the account illuminates how the 'explains' part works. This debt is paid by appeal to an analogy with the operation of a computer program, where the implementation of the program requires that lower-order electrical properties are brought into play. So the explanation is secured in similar fashion to the way that appeal to causes secures explanation: the effect is *brought about* by the explaining feature(s). This example makes it look like programming properties secure the causal effectiveness of the lower-order properties they program for in a causal way. But it is clear that this is not how programming properties are envisaged as carrying out their programming work; they are meant to carry out this work in a non-causal way. That is the point of the PE strategy as applied to the problem of mental causation – that it avoids over-determination problems by taking physical properties to carry out the causal work and taking mental properties to *non*-causally ensure that there will be physical properties available to carry out that work.

As we see it, there are two, related, problems that make this solution implausible. The first surfaces when one asks how in general such programming is effected. How, exactly, do the higher-order properties non-causally *ensure* that the causally effective lower-order properties are instantiated? The second appears when one considers the most attractive answer to this question.

Consider the first. The problem is that, on a plausible reading of Pettit's account, the higher-order and lower-order properties are separately instantiated, and

---

[7] Pettit mentions two other examples, that of boiling water causing a crack in its container, and the failure of a square peg to fit into a round hole. Neither of these are cases of higher-order properties programming for the instantiation of their realizing properties, so we set them aside.

the most natural way to interpret the *ensuring* relation between *different* property-instances is causally. Clearly this interpretation of *ensuring* depends upon the property-instances being separate (non-identical) so we need to support this understanding of what Pettit says.

There are two reasons for thinking that the program model requires that the property-instances be non-identical. The first derives from how Pettit describes its essential features. Take the three conditions he places on a property's being a 'programming property':

> 1. Any instantiation of the higher-order property non-causally involves the instantiation of certain properties – maybe these, maybe those – at a lower order.
> 2. The lower-order properties associated with instantiations of the higher-order, or at least most of them, are such as generally to produce an E-type event in the given circumstances.
> 3. The lower-order properties associated with the actual instantiation of the higher-order property do in fact produce E. (Pettit 1993: 37)

What these quoted passages make clear is that Pettit's model works not only with an ontology of higher-order and lower-order properties, but also of higher-order and lower-order 'states', or instantiations of properties.[8] It is also clear that if the proper construal of the relation between the instances were that they are identical, it would have been easy to say so, rather than to leave it at the vague and non-defined 'association' between the instances. So we think that a natural reading of these passages forces the non-identity interpretation, and with it the thought that the *ensuring* relation between the separate instances is a causal relation, it being plausible to think that one event (property-instantiation) ensures that another occurs only if it causes, or causally contributes to, the occurrence of the latter.

The temptation to interpret *ensure* causally is strengthened by another locution employed by Pettit: "The factor that programs for an effect…non-causally *arranges things* (it means that things are arranged) so that there will be such a producer state –

[8] There is a question of how this can be reconciled with a further view, articulated a few pages earlier in his account of causal relevance, that there is token identity between the role (higher-order) state and

maybe this, maybe that – available to do the work." (Pettit 1993: 37, italics added)[9] How one property can be such that its instantiation non-causally arranges things so that an instance of another property produces the effect, though, is mysterious. What seems to be required here is a *de re* relation between separate instances of two properties such that the one instance (non-causally) produces (ensures, arranges for) the second instance, this latter instance producing (causing) the effect.[10] The natural reading of *both* 'producing' relations here is causal, and we have been given no reason to think otherwise, apart from the convenient fact that making the first production non-causal avoids causal over-determination. So on our favoured reading the ensuring relation holds between (separate) property-instances, and this makes it difficult to see this relation in anything other than causal terms.

Our second reason for the 'separate-instance' interpretation arises from considering a different understanding of what is going on, one that says: the 'arrangement' of the higher-order and lower-order *properties* is such that whenever the one is instanced, so is the other. Here the ensuring work is done at the property level, so the accusation that causality is implicated is neatly avoided. And this reading is supported by a functionalist account of the relation between mental and physical properties, a relation of role property to realizing property, where a mental state (such as pain, or the belief that *p*) is the occupant of certain causal role, the causal role definitive of the mental type in question (in the Ramsey/Lewis style). Now, there is no difficulty understanding the claim that the role properties non-causally ensure the existence of realizing properties, but the most natural way of making metaphysical sense of this connection in these cases is by *identifying* the instantiations, and, for reasons to which we return later, that is a result Pettit wishes to avoid. For one thing, this result would also ensure that the higher-order property is causally efficacious,

---

the realizer (lower-order) state (Pettit 1993: 33) (though we note that Pettit, rightly in our view, does not think this solves the problem of the causal relevance of properties).

[9] Again, by 'producer state' here Pettit seems to mean, not 'property' but 'property-instance'.

[10] Actually, this makes it look as though Pettit might be flirting with a trope conception of properties, rather than a universalist one. Given that, on the latter view, an instance of a property just is the thing that has it, it is difficult to make sense of what is going on here without supposing either that the mental event and the physical event are distinct (in being distinct instances of distinct properties), which Pettit clearly is not supposing, or that the property-instances, in being distinct instances, are tropes. This latter would allow him to say that mental events are physical events, but only by committing him to the view that such events are constituted by both mental tropes and physical tropes. Trope theories of mental causation also suffer from the problem of causal relevance, and we think they are in a worse position than the PEA to resolve it, though we haven't the space to go into the reasons why here. We discuss them in detail in C. Macdonald and G. Macdonald (forthcoming).

contradicting the claims made by advocates of Program Explanation.[11] For another, this co-instantiation account is one we have defended elsewhere (and below), and it has been criticized by Pettit for having the defect that "it would make a state like the belief that *p* causally relevant but relevant in virtue of a property other than that of being the belief that *p*: relevant in virtue of being such and such a neural or electronic state." (Pettit 1993: 38) Although this criticism is couched in terms of causal relevance rather than causal efficacy, it is clear that the identity claim is being rejected, and with it any chance of rescuing the identity of the higher-order and lower-order property instances.[12] We are left with the puzzle as to how the higher-order instance can non-causally ensure the instantiation of the lower-order property.

Friends of the PE strategy might say that we are ignoring Pettit's explanation of how programming works in the case of mental causation. Consider an agent's intentional state, say, *S*'s belief that *p*, which explains an agent's action of *A*-ing. The intentional state will program for the *A*-ing if, no matter how it is realized (variable realization being possible), the realizer (lower-order, physical) state, say, a particular neural state, tends to produce a type-*A* action. (By 'state' Pettit here must mean 'property' rather than 'property-instance', given that he takes the realization relation to hold between properties. See note 6.) That is, all of the possible realizers of the belief that *p* must be such that they tend to produce *A* (given certain background conditions). The intentional state, the belief that *p*, is causally relevant to the agent's *A*-ing just because it is realized by states, all of which tend to cause *A*-type actions (given the background conditions). But this fact, that the intentional state is realized only by producers of *A*-type actions, cannot be just a happy accident, merely fortuitous. If it were, then all cases of mental causation would depend on the huge coincidence that relevant higher-order intentional states are all realized, it so happens, by states with an appropriate causal profile. To reduce such coincidence, Pettit relies on the assumption that agents are *designed* so that their higher-order intentional states will be realized by the appropriate lower-order, causally efficacious, neural states. This design assumption underwrites the 'program' part of the PE model: given

[11] The claim that the identity of instances guarantees the causal efficacy of the higher-order property is defended below, but was first proposed by us in C. Macdonald and G. Macdonald (1986). Note that this does not guarantee the causal *relevance* of the higher-order property, but it is not *relevance* that is at issue here. See Section 5 below.

[12] We return to consider this objection, particularly as an objection to an account of causal relevance, at the beginning of Section 5.

adequate design, the instance of the higher-order property will ensure that there will be an instance of the lower-order property.

Does this help with the problem of how the higher-order instance non-causally arranges for there to be an instance of the lower-order property? We don't think so. On the contrary, we think it makes the non-causal aspect of the story more implausible. The natural way to think of design in this context is biological design, and this is the way Pettit thinks of it: "…we may readily assume that any natural intelligence is going to have been designed, under evolutionary and perhaps cultural pressures, to meet suitable design specifications." (Pettit 1993: 41) One must presume that the designer here is natural selection, where such selection includes cultural selection, whatever that turns out to be. But any selection-style story about design will be resolutely causal, relying on the selection process operating on the effects of the properties it selects. Brutally briefly, the story will be that some properties have instances whose effects in a particular environment make their possessors more likely to replicate themselves (more likely than competitors lacking the relevant properties), thus producing more instances of those properties. This clearly requires that the property-instances be causally efficacious in the selection process, so this part of the PE model strengthens the feeling that the causal power of the intentional property is being discarded in an unprincipled way, just to avoid the over-determination problem.

A critic may point to cases in which a property is 'non-causally' selected because it is regularly correlated with a property for which one can tell the appropriate causal story. But reliance on such regular correlations is not available to the PE model, as it is trying to *explain* such regular correlations between instances of higher-order properties and instances of appropriate lower-order realizing properties.

This brings us to the second, related problem with the PE model mentioned earlier, which is that many accounts of causality would have trouble denying the status of cause to the instance of the higher-order property. Consider, for example, a counterfactual account. Since the PE model requires that, had the higher-order property not been instanced, the effect would not have occurred, such an account renders the higher-order property a cause of the effect. Or, consider Woodward's recent 'manipulation theory' of cause and causal explanation, which states that "as a rough approximation, a necessary and sufficient condition for X to cause Y or to figure in a causal explanation of Y is that the value of Y would change under some

intervention on X in some background circumstances…." (Woodward 2003: 15). This also makes an instance of a higher-order property a cause. In order for the PE strategy to work in the requisite way on this theory, it is necessary for the supposedly non-causally productive instance to be susceptible to interventions that would change instantiations of the lower-order property, and hence change the effect produced. But if so, then Woodward's condition on something's being a cause is met. Again, apart from the convenience of avoiding a problem (over-determination), it is difficult to see why one would want to deny that the instance of the higher-order property is causally efficacious.

4. Our Alternative

Our diagnosis of what goes wrong with the PE model is that it assumes that the higher-order and lower-order properties must have *distinct* (non-identical) instances. This is what generates the puzzle about how it is that the first instance non-causally produces the second instance. Our view is that one can employ the same metaphysics of properties and instances (presuming a universalist, rather than a tropist, view of properties) and rescue the causal efficacy of the mental by rejecting this assumption, i.e., by identifying the instances of the mental and physical properties of events (as well as by identifying the events that exemplify them).

A more comprehensive metaphysical basis for this resolution is supplied by the Property-Exemplification Account of events, (PEA).[13] As noted earlier in our discussion, according to it, events are exemplyings of (*n*-adic) act-or event properties at (or during intervals of) times in objects. The objects in which such exemplifyings occur are the subjects of those events. And the properties, whose exemplifyings in subjects just are events, are properties, not of events, but of their subjects. For example, the event of Jones' running at noon yesterday just is the exemplifying in Jones of a property of Jones, the property, *runs*, at noon yesterday. Such properties are termed constitutive properties of events, and are so termed because they are the properties of subjects whose exemplifyings by those subjects just are events. Constitutive properties of events are properties whose exemplifying it is of the essence of those events to be.

In addition to constitutive properties, events also have characterizing properties. These are properties events possess, at least some of which they possess in virtue of 'having' (i.e., being exemplifyings of) constitutive properties. Thus, for example, the event that is the exemplifying of the property, *runs*, by Jones at noon yesterday, has as its constitutive property a property of Jones. That event has the property of being a running.

Events construed along these lines are sometimes referred to as 'structured particulars'. They are deemed so because they 'have' not only constitutive properties, but also constitutive objects (or subjects) and constitutive times.[14] That is to say, it is in the nature of any event to be an exemplifying of a property (of its subject) in a subject at a time. Two conditions on events are essential to the account, one an existence condition and one an identity condition. These are formulated for monadic events as follows:

> Existence Condition: Event [*x*,*P*,*t*] exists if and only if the object *x* has the property *P* at time *t*.

> Identity Condition: Event [*x*,*P*,*t*] is identical with event [*y*,*Q*,*t'*] if and only if the object *x* is identical with the object *y*, the property *P* is identical with the property *Q*, and the time *t* is identical with the time *t'*.

---

[13] We supply the necessary metaphysical details to allay the doubts of those who may think it cannot be supplied. One such person is Stephen Yablo (1992), esp. p. 259.

[14] The exposition of the PEA here is based on work of Kim's (esp. Kim (1976)). According to Kim, although the first condition on events specified here is indispensable to the theory, the second, as formulated, is not. The theory could proceed, for example, by defining the predicate 'is an event' over ordered *n*-tuples of objects, properties, and times. In this case, the ordered triple, <*x*,*P*,*t*>, would be an event if and only if *x* has *P* at *t*; and the principles of set theory would guarantee the existence of the triple (assuming, of course, that *x*, *P*, and *t* exist). But Kim himself appears to favour the first method over the second, and it is certainly the preferable one from the point of view of the phenomenon of causal interaction between events, where this is assumed to entail their positionality. The claim that events have constitutive objects, properties, and times, should not be understood as the claim that they are in some way constituted by or composed of objects, properties, and times, these being related to each other in something like the way that a chair, say, is often viewed as composed of or constituted by its parts arranged in a certain way. This much is clear from the fact that the relationships that the 'components' of events bear to one another are very different from the relations that the components of physical things bear to one another. In the case of an event, one component is *exemplified by* another, *at* yet another; and it is clear that whatever the constituents of a biological organism or an artifact may be, they do not bear this relationship to one another. Given all of this, the claim that the components of events are constitutive of them amounts to the claim that they are essential to them. Kim explicitly commits himself to some version of the latter. For more on this, see Macdonald (1989). Lombard (1986) agrees with Kim that the identity condition on events, as formulated, is not essential to the account, but for different reasons. According to him, events can have more than one constitutive property, whereas the identity condition as formulated here assumes that each event has just one.

where *x* and *y*, *P* and *Q*, and *t* and *t'* are variables ranging over objects, properties, and times, respectively.

The PEA construes properties as both abstract and multiply exemplifiable; entities that can have, but are not identical with, their exemplifyings. According to it, to say that a mental event is identical with a physical event is to say that each event which is (= is identical with) an exemplifying of a mental property of a subject in that subject at a time is identical with an exemplifying of a physical property of that subject in that subject at that time. So, to say that a mental event is a physical event is to say that there is just one exemplifying of two properties, one mental, and one physical, by an object at a time. Thus, appealing to the PEA in order to rescue causal efficacy for mental events requires simply recognizing that a single event can be (identical with) an exemplifying of both a mental and a physical property.

What is the relation between mental and physical properties of persons – those properties whose exemplifyings just are mental/physical events? On our view, mental properties whose exemplifyings just are mental events are not constitutive properties of those events, but rather, supervene on physical properties constitutive of such events, and, consequently, mental properties of events supervene on physical properties of events.[15] Given the identity condition on events imposed by the PEA, non-reductive physicalism requires rejection of the view that mental properties are constitutive properties of the events that have them, on the assumption that each event has only one constitutive property (but see note 14). But, independently of this, the position is committed to some kind of supervenience thesis, since without such commitment it is difficult to fend off the charge that the position is irredeemably dualist because it acknowledges the presence in the natural world, if not of non-physical events, of non-physical properties. Many will think that such a position does not deserve the name 'physicalist'.

What kind of supervenience thesis best captures the relation between mental and physical and physical properties is a thorny issue, as is well known.[16] Still, for

---

[15] Does the fact that mental properties are not constitutive properties of events show that physical events that are mental events are not in some sense 'genuinely' mental? No, only that physicalism is contingently true.

[16] For some skeptical discussion of the value of appeal to psychophysical supervenience, see, for example, R. Miller (1990), A. Melnyk (1995), and J. Heil (1998). For some examples of work on psychophysical supervenience that seeks to meet objections based on the claim that no satisfactory thesis can be found, see T. Horgan (1993) and T. Grimes (1991).

present purposes we can say this much. Take supervenience between the mental and the physical to be that relation which holds between a mental property or set of properties, *M* and another, physical one, *P*, such that any two objects/events indiscernible with respect to *P* cannot diverge with respect to *M*. Further, following Kim (1978, 1984), let us distinguish weak from strong supervenience. Then we can define a relation of strong supervenience thus:

> SS: *M*-properties strongly supervene on *P*-properties =*df*. For any possible worlds *w* and *w**, and any individuals *x* and *y*, if *x* in *w* is a *P*-twin of *y* in *w**, and the actual world's laws of physics hold in both, then *x* in *w* is an *M*-twin of *y* in *w**.[17]

where any *x* and *y* are *M* (/*P*) twins if and only if *x* and *y* are exactly alike with respect to their *M* (/*P*) properties.

What does the issue of causal relevance of mental properties amount to in this context - the issue of whether the mental property *of* an event is causally effective in that event's bringing about the effects it does?[18] Well, according to the PEA, events have (characterizing) properties as well as being the exemplifyings of properties. They have properties by exemplifying them. Given this, and given the universalist understanding of properties to which PEA subscribes (see note 4), whereby an instance of a property just is (i.e., is identical with) the thing that has (exemplifies) it,

---

[17] This is an adaptation of the definition of strong supervenience given by Brian McLaughlin in McLaughlin (1995). By *M*-properties (/*P*-properties) we mean the non-empty set, *M* (/*P*), of properties. We choose this version over Kim's principally because it is weaker than Kim's, though Kim's entails it. Kim's implies that it is necessarily the case that if something has an *M* property, then it has some *P* property. But SS could be true if twins had no *P* property at all. It thus allows for the possibility that there might be purely mental worlds. We think this consequence desirable, given that we take physicalism to be true and contingent, and given the possibility of variable realization of mental properties. Kim blocks it only by assuming that *P* properties are properties of the *P*-type and by allowing complementation to be a property forming operator, so that *–P*, a negative property, is a way of being of the *P*- type. But one might deny that there are any negative properties on the grounds that they have no causal powers, and we think that this is a plausible thing to do. We adapt McLaughlin's definition primarily by introducing the caveat 'and the actual world's laws of physics hold in both' precisely because we take physicalism to be true and contingent. Although we do not take the identity conditions of properties to be given in terms of their causal powers, we do take it that if properties are identical they have the same causal powers, and we take it that physical laws relate properties in virtue of their causal powers. So worlds in which there are the same physical properties that there are in the actual world will be worlds in which the same physical laws hold.

[18] Again, as we point out in the text, we take the causal relevance of a property to be more than a matter of its instances being causally effective, i.e., causally efficacious. But we do take this to be a necessary condition on causal relevance, so for present purposes we will not allude to other conditions that we

instances of mental properties of mental events are identical with instances of physical properties of physical events (since each mental event is identical with a physical event). We take it that a necessary (but not sufficient) condition on causal relevance of properties is that instances of those properties are causally efficacious.[19] So, to say that a mental property *of* a physical event is causally relevant is at least to say that an instance of that property, i.e., that event, is causally efficacious in bringing about an effect of that event. According to our strategy, this will require that (mental) instance to be a physical instance, i.e., that one and the same event is an instance of both a mental and a physical property.

That requirement would need to be met anyway, since, on the universalist conception presumed by the PEA, things exemplify properties, and a thing just is (i.e., is identical with) an instance of each property that it has. Thus, an event exemplifies its properties, and it is (= is identical with) an instance of each property it has. This alone makes it difficult to see what exactly is going on in the metaphysics of Pettit's PE strategy, since he takes mental events to be identical with physical events, but wants to distinguish instances of higher-order properties (of events) from instances of lower-order ones, apparently using the PEA. We think this isn't possible. So, given the metaphysics of the PEA and Pettit's commitment to psychophysical event identity, there is independent reason to reject the 'distinct property-instances' view.

But we also hold a further thesis, which applies specifically to higher-order and lower-order properties. Things – objects, events, and other individual particulars – exemplify properties, but some properties they exemplify just by exemplifying others. This is not the case with all properties that a thing exemplifies, even given the universalist conception of properties. But it does apply to some. Consider a red, square box. It has the properties of being red and being square. It also has the property of being coloured. It is (identical with) an instance of each property that it has. The box exemplifies the properties of being red, being square, and being

---

take to be necessary as well. For more on the other conditions, see note 19 below and C. Macdonald and G. Macdonald (1995).

[19] It is clearly not sufficient, since, given the view that an exemplification of a property is the thing that has it, every property exemplified by an event would be causally relevant whenever that event caused any effect. So, in addition to (1) causal efficacy of their instances (i.e., events) we place two further conditions on the causal relevance of properties: (2) that the properties participate in a general 'pattern', or network of relations, in nature (one example of which is the nomological pattern), and (3) the generality which the properties display must be of the right type for a given type of effect (e.g., if the pattern is the nomological pattern, the nomological property must be nomological for a certain type

coloured. It does not exemplify the property of being square just by exemplifying the property of being red. But it does exemplify the property of being coloured just by exemplifying the property of being red.[20]

Similarly, events exemplify properties. But some properties they exemplify just by exemplifying others. In the case of higher-order (mental) and lower-order (physical) properties of events, we claim that this is just what happens. Thus, a mental event can exemplify the property, *being a thinking of Vienna*, just by exemplifying the property, say, *being neuro-chemical event* α. We claim that this view has an independent plausibility; it is not invoked just to solve a problem of over-determination. Where one property (or properties) of an event is said to *realize* another property (or properties) of that event, this is by far the most plausible way to construe the relation between the properties exemplified.

Call this thesis the Property-Dependence Thesis, to distinguish it from another, weaker thesis that we also hold and which follows from the universalist conception of properties, the Co-instantiation Thesis.[21] According to the latter, a mental property and a physical property of an event can be co-instantiated in a single instance, i.e., there is just one instance of two properties. This thesis is weaker than the Property-Dependence one, because on the universalist conception, *all* properties of an event are co-instantiated in a single instance – an event is just one instance of all of its properties, not just the higher-order and the lower-order ones. Still, since, on our view, the stronger thesis entails the weaker one, it follows that, where P is the physical property realizing mental property M, there will be just one instance of both P and M, P_i, i.e., M_i.[22] Given this, the mental instance, M_i, will be causally efficacious whenever the physical instance, P_i, is – given the assumption we share with Pettit, that causes are worldly events.

---

of effect). For more on these conditions and our defense of them, see C. Macdonald and G. Macdonald (1995).

[20] But note that we do not think that the relation between mental and physical properties is a determinable/determinate relation, as some others (e.g., Yablo 1992) do. So, in the case of the properties of being coloured and being red, the dependence thesis holds because that the object is an instance of the property of being colored is entailed by the fact that it is an instance of the property of being red. In other cases, of which the mental/physical property case is one, the dependence will hold for a different reason.

[21] See C. Macdonald and G. Macdonald (1986).

[22] So, in what follows, when we claim that $P_i=M_i$ – that is, that there is one instance of both a mental property and a physical property where mental properties are not identical with, or reducible to, physical ones – we mean more than just that there is just one instance of both a mental property and a physical property (since on the universalist conception of properties, an event is just one instance of *all* of its properties). We mean that a mental event exemplifies M just by exemplifying P.

In short, our claim is twofold: (1) mental properties of persons supervene on their physical properties, and (2) mental properties of events supervene on their physical properties. This is consistent with the view that an individual event can be an exemplifying of both a mental and a physical property (of a person), can be an instance of both a mental property and a physical property (of an event), and can be an instance of a mental property just by being an instance of a physical property (of an event). Consequently, the epiphenomenalism problem that attaches to mental properties disappears, along with causal over-determination.

5. Resistance to the Preferred Alternative

The temptation to resist the suggestion that there is just one instance of two properties, one mental and one physical, is, we think, a result of two separate but related thoughts, both of which we think are mistaken. The first is the thought that distinct (non-identical) properties cannot share their instances, cannot be co-instantiated. But this thought is clearly wrong. Using an analogy we have used before, whenever the property *being red* is instanced, so is the property *being coloured*; it is very implausible to view the situation as anything other than one of a single instance of *both* properties, despite the non-identity of the properties being instanced. In general, in any case of properties related as determinate to determinable, an instance of a determinate property will just be an instance of a determinable property.[23] The property of weighing less than 100 lbs can be instanced by instancing the property of weighing 2 lbs. The latter will also be an instance the properties of weighing less than 99 lbs, of weighing less than 98 lbs, and so on. (Note that these are also all examples of cases in which the stronger, property-dependence thesis is true.) It seems to us to be ontologically promiscuous to populate the world with extra instances whenever this happens. There is just one instance, which happens to be of many properties.

We stress that this is an analogy only: others have viewed it as no mere analogy, and have modeled the relation of mental to physical properties as a relation between determinable and determinate properties.[24] We do not think this is a correct

[23] We first suggested this analogy in C. Macdonald and G. Macdonald (1986). For further elaboration see C. Macdonald and G. Macdonald (1995).
[24] See especially Yablo (1992).

analysis of the mental-physical relation. One cannot infer from the fact that a person has a brain-property α that they have a mental property β (at least not without considerable empirical theoretical input), so one cannot infer from the fact that an event (an exemplifying of the brain-property α, i.e., an exemplifying of the mental property β) has the property of being an instance of the brain-property α that it has the property of being an instance of the mental property β. However, one *can* infer from the fact that something has the property of being red that it has the property of being coloured. Still, the determinate-determinable relation is just one example of a supervenience relation, and there are other examples of this latter relation that can underwrite our confidence that mental properties are co-instantiated with the physical properties upon which they supervene. In particular, that higher-order properties are *realized by* lower-order properties makes the identification of their instances compelling. At the very least the onus is on those who deny this identification to explain how the realizing instances come to be 'separate existences'.

The second thought responsible for resistance to our solution is that, although co-instantiation is possible, it does not provide for the causal relevance of the higher-order property. As we have seen, Pettit expresses the objection that on our view intentional states are given relevance "through construing them as identical with electronic or neural states", this having the defect that "it would make a state like the belief that *p* causally relevant but relevant in virtue of a property other than that of being the belief that *p*: relevant in virtue of being such and such a neural or electronic state." (Pettit 1993: 38)[25] We take it that this objection has the following form: being the belief that *p* will be causally relevant when it is capable of explaining an effect that would have been produced by instances of *any* of its realizing properties. Because its relevance is thus 'general' with respect to all of its realizers, the causal relevance of the higher-order mental property cannot be identified with the causal relevance of any particular (lower-order, physical) realizing property. But, so the objection goes, our account, in identifying the *events* as it does, identifies the causal relevance of the mental property of the event with the relevance of a particular physical realizer property of that event.

[25] We cite Pettit here because of the context, but this objection is common to several critics of our view; see Ehring (1996, 1999) and Yablo (1992), for example. Note that Pettit's use of 'state' here must mean 'property-instance'.

It is worth clarifying what would be wrong with such an identification from our point of view. One could only identify the causal relevance of an intentional property (say, the property of being the belief that *p*) with that of a particular realizer of that property (say, the property of being neuro-chemical state α) if *that* realizer's relevance was shared by all the other (possible) realizers of the intentional state. This identification, if correct, would lead immediately to the reduction of the intentional property to its realizer, there being no other realizers whose causal relevance differed, thus defeating our aim of establishing the possibility of non-reductive monism.

What the criticism does, though, is conflate the requirements on causal efficacy, which concerns events, with those on causal explanation, which concerns what we have been calling causal relevance. As noted above, our view is that the causal efficacy of events is just one condition on the causal relevance of the properties of those events. The notion of 'being relevant *in virtue of* property *P* rather than property *Q*', whether this concerns properties that are co-instanced or whether it concerns properties connected by the 'ensuring' relation, can only be made sense of in terms of the causal power of the property instanced, this power making that property relevant for the causal explanation of (an aspect of) the effect. The property, *being the desire that q*, will be relevant to the causing of a number of appropriate actions, appropriate given the content of the desire. Those actions will be caused, in part at least, *in virtue of* the agent's having that desire (i.e., exemplifying the property, *desires that q*). On different occasions a desire with the same content will be variously realized (given non-reduction) by different first-order physical properties, and while those realizers may be relevant to (and so explanatory of) the actions physically described, the claim is that it will be *in virtue of* the desire's being a desire that *q* that the action, intentionally described, is performed.

If this is what is meant by *in virtue of*, and the related notion of causal relevance, then we agree with Pettit that it is only some properties of events that are causally relevant to the effects they produce. Which properties they are will depend on which aspects (properties of) the effect one wants explained. But none of this touches on the point about causal efficacy. It is clear that on our account not all properties whose instances are causally efficacious will be causally and so explanatorily relevant (though we do insist that any causally relevant property must be one whose instances are causally efficacious). Thus, suppose that a window has a shattering point of 5 lbs, and shatters because it is struck by a 7 lb rock thrown at it. It

is causally relevant that the rock weighs over 5 lbs, even though this instance of weighing over 5 lbs is also an instance of weighing over 2 lbs. The throwing of a rock weighing over 2 lbs won't explain the window's shattering, though *in this case* the instance of the property, *weighing over 2 lbs*, just is an instance of weighing over 5 lbs (= an instance of weighing 7 lbs), and so the throwing caused the shattering. The property, *weighing over 2 lbs*, does not help explain the shattering because other instances of it in rock-throwings won't be ones in which the window will shatter. But this truth about other instances is irrelevant to the causal efficacy of this instance, it being pertinent only to the matter of causal relevance.

It is difficult to see why this outcome should be problematic. *Any* view of the causal relation that takes it to be an extensional relation, relating items 'in the world,' will have the consequence that some causally efficacious properties, properties whose instances are causally efficacious, will not illuminatingly explain the effects they bring about. They will not, in these cases, be causally relevant properties. The only way to avoid this result is to drastically change the metaphysics presupposed here, that of properties and instances and the PEA account of events, a metaphysics to which Pettit seems committed.[26] Short of this, one can only avoid this consequence (on the assumption of extensionality) by insisting that no causally efficacious instance can be an instance of more than one property, thus ensuring that causal efficacy and relevance cannot come apart. This is an extraordinarily strong claim to make given the obvious counterexamples presented by instances of different determinates of a determinable. Take the example just given. The rock's weighing over 5 lbs is relevant to (and so explanatory of) the window's shattering; its weighing over 2 lbs may be relevant to my toe's hurting when it falls onto that toe. Its weighing 7 lbs is both its weighing over 5 lbs and its weighing over 2 lbs; one surely multiplies instances beyond necessity by insisting that these are all distinct instances.

Pettit himself needs something like the distinction between efficacy and relevance, even on his account where programming properties are said to be causally relevant but not causally efficacious. On this account, some properties that program for their effects do not figure in illuminating accounts of why those effects occurred. That the eraser was elastic programs for its bending, but its elasticity can be inferred

[26] We explore, and reject, a challenge to this metaphysics stemming from a specific version of trope theory (Robb, 1997, 2001), which takes the properties whose causal relevance is in question to be tropes themselves, in C. Macdonald and G. Macdonald (forthcoming).

from knowledge of its molecular structure, making the dispositional property (elasticity) 'insignificant'. So, Pettit adds the requirement that the programming property be 'significant', thus ensuring its explanatory relevance. If causal relevance requires 'significance', we can avail ourselves of this resource without going the circuitous and difficult route of 'programming'. We say that only some properties whose instances are causally efficacious are 'significant', or causally relevant, to their effects.[27] That some are not is irrelevant as an objection to our account.

It is worth noting that in the example just cited, that of elasticity, it would be strongly counter-intuitive to insist that when the dispositional property of being elastic is realized, its instances are not instances of the realizing property. In general, we claim, whenever a higher-order property is a functional property, it is co-instanced with its realizing properties – and so is as causally efficacious as they are. That being elastic is not a *significant* programming property is not germane here, since that does not change its status as a programming property. On the PE model, what lack of significance does is render the property causally irrelevant and so non-explanatory. This might make it look as though no higher-order property *can be* causally relevant, in the sense that it can have a causal profile that cannot be identified that of their lower-order realizers. But we deny this too, for the reasons given in section 4 above.

We take it that the above establishes our strategy as intuitively more acceptable than one that makes mental properties causally inert. It secures the causal efficacy of such properties, and avoids over-determination, all in an ontologically parsimonious way. Moreover, its main device, co-instantiation, receives independent support from considerations stemming from the realization relation. However, its claim to be the best available strategy has been challenged (Kim 1999) and this challenge could be seen to provide indirect support for Pettit's view that the higher-order properties are not causally effective in producing the events that they explain. Kim's challenge takes the form of an argument that purports to show that if supervenient properties have causal powers then they are reducible to the properties on which they supervene. If correct, this would defeat our aim to show how

[27] The conditions for causal relevance are spelled out in more detail in C. Macdonald and G. Macdonald (1995). We think that the problem of 'significance' will be inevitable on any account of causal relevance, since relevance has to do with explanatory potential, and this will vary from context to context, depending on the type of effect to be explained. Moreover, even within a single context, there will be properties whose instances are causally efficacious but the properties themselves will not be causally relevant, as the above examples illustrate. This just goes to show that it is a mistake to think that there is such a thing as causal relevance *tout court*.

irreducible mental properties can be both causally efficacious *and* causally relevant, since, in order for our solution to work, it must be possible for there to be emergent properties. We defend this possibility below, again on the basis of the crucial distinction between properties and their instances.

6. The Possibility of Causal Relevance

For some time now Jaegwon Kim has been arguing that the non-reductive monist's picture of the mind is seriously unstable, the instability making the 'non-reductive' aspect untenable. Consider Fodor's non-reductive monism, where the reduction of mental properties is rejected on the grounds that they are variously realized, where this means that their base realizing properties are heterogeneous. Given this heterogeneity, it is claimed that the subvening property formed by disjoining the particular realizing properties will not form a natural kind, thus blocking reduction to that disjunctive property. The lower-order properties cannot form disjunctive antecedents and consequents of a single law, so a bridge law is ruled out.[28] The problem, as Kim sees it here, is that we need an answer to the question of why the supervening property, say pain, is not "…equally heterogeneous and nonnomic as a kind" (Kim 1993: 323). Failing one, we must take seriously the thought that the variably realized supervening properties cannot figure in laws, not even *ceteris paribus* laws, leading to the conclusion that there are no 'special sciences' (on the assumption that all scientific explanation is law-based).

A related argument concentrates on the causal powers of the supervening and subvening properties. This begins by noting that the non-reductive monist is committed to there being a difference between the causal powers of properties at the subvening and supervening level, those at the subvening level forming a heterogeneous set, those at the supervening level supposedly being homogeneous. The threat that arises is that the supervening properties will have no unified causal powers, thus making them causally irrelevant.[29] Recently Kim has advanced an argument along these lines, coming to a conclusion that has the form of a dilemma:

---

[28] Thus, he says, "the lower level disjunctive antecedent is not a natural kind, and so is not law-apt – "…a badly heterogeneous disjunction is unsuited for laws" (Kim 1993: 318). Note that Kim's most recent view (Kim 1998, 2005) is that bridge laws are neither necessary nor sufficient for reduction, since there is functional reduction (which he endorses).

[29] For an early formulation of such an argument see G. Macdonald (1986).

either the supervening mental properties are causally inert, or they are reducible (Kim 1999, 2003). Either way, it is bad news for non-reductive monism, which needs irreducible higher order properties that possess (independent) causal powers.

Such properties would be, in our terminology, emergent ones, having a distinctive causal profile. According to Kim's argument, if there were such emergent properties then 'downward causation' would be possible, and downward causation is incoherent. The argument for downward causation goes like this. Emergent properties must have distinctive causal powers. They must be capable of being causally effective in bringing about their own distinctive effects. Suppose that they only bring about effects of the same (higher-order) level. These effects will be higher-order effects (given that emergent properties themselves are higher-order). But this means that the higher-order effects will have lower-order realizations. So, it is by causing instances of the lower-order (base) realizing properties that an emergent property will cause a higher-order effect. So higher-order causation presupposes downward causation.

Why, according to Kim, is downward causation incoherent? Consider emergent properties *M1* and *M2*, where *M1* causes *M2*'s instantiation, *M1* being realized by *P1* and *M2* realized by *P2*. Given that *M2* 'arises out of' (is realized by) *P2*, *M2* would be instantiated by *P2*'s instantiation, regardless of whether *M1* had caused *M2*. Simplicity dictates that *M1* causes *M2*'s instantiation by causing *P2* to be instantiated (the 'Downward Causation' conclusion). But g*iven* that *M1* is realized by *P1*, and *given* irreducibility (i.e., that $M1 \neq P1$), we now have two sufficient causes of *P2*. This embarrassment of causal power can only be resolved by (a) eliminating *M1* or *P1* as a cause of *P2*, or (b) sacrificing irreducibility. Pettit takes the first option, giving up on the causal power of the mental, Kim the second.

Kim concludes that the emergent property *M1* does not independently cause *P2*'s instantiation: what is doing the causal work is what realizes *M1*, namely, *P1*. So the so-called emergent property has no (distinctive) causal power, and *M1* has no independent causal relevance. From this, given the previously stated assumptions, we can conclude that the special sciences cannot be defended by relying on the model that takes special science properties to be higher-order properties that are variably realized by, but irreducible to, physical ones. "If emergent properties exist, they are

causally, and hence explanatorily, inert and therefore largely useless for the purpose of causal/explanatory theories" (Kim 1999: 33).[30]

7. A Different Argument Against Downward Causation

Given the hierarchical picture of the sciences presented here, it might look as though Kim has a sound argument for the causal irrelevance of emergent properties. But we claim that the argument is not sound. In this final section we want to show that there is a *sense* in which it is true that downward causation is incoherent. But the route to that conclusion is significantly different from Kim's, and leads to different consequences. In particular, it rescues the possibility of the causal relevance of (some) higher-order properties, mental ones included.

The argument, as presented, shuttles between talking of the downward causal power of properties and that of their instances. It is not that Kim is unaware of the importance of the property/instance distinction. He recognizes that "Properties as such don't enter into causal relations; when we say *M* causes *M**, that is short for 'An instance of *M* causes an instance of *M**' or 'An instantiation of *M* causes *M** to instantiate on that occasion'." (Kim 2003: 155)[31] But if we keep this distinction in mind, his conclusion that 'higher-order causation is downward causation' does not follow as immediately as he thinks it does. The crucial move in the argument is taken when downward causation is said to be required even for causation at the same (higher-order) level. The higher-order 'effect' (*M2*) is realized in a lower order property (*P2*), and it is *an instance of* the lower-order property that is caused by (*an instance of*) the higher-order *M1*. As noted above, *M1* does this by being realized by *P1*, the consequence being, so Kim argues, that either *P1* does all the causal work, or *M1=P1*. Kim opts for the latter solution, rescuing the *M1–M2* 'causal' relation by ensuring, via reducibility, that it is the same relation as the *P1-P2* 'causal' relation.[32]

---

[30] This argument is updated in Kim (2003). In the earlier paper Kim took the argument to show that the mental was causally inert, in the later paper he stresses the reducibility of the mental. If we can show that the mental can be causally relevant we will have defused the argument for reducibility.

[31] He also says: "The fact that properties *M* and *P* must be implicated in the identity, or non-identity, of *M* and *P* instances can be seen from the fact that "An *M*-instance causes a *P*-instance" must be understood with the proviso "in virtue of the former being an instance of *M* and the latter being an instance of *P*"…" (Kim 2003:157). On our view, this conflates causal efficacy with causal relevance.

[32] Our use of scare quotes around key terms here in this paragraph is intended to mark the equivocation we detect in the argument between talk of property-instances and causal efficacy, and talk of properties and causal relevance.

Diagramically, his picture of the situation is this (Kim 2003: 166):

M1 -------causes---→ *M2*
= is reductively =
identical with
P1 -------causes-----→*P2*

Figure A

i.e., causation between mental properties just is causation between physical properties, since mental properties are physical properties. But this picture plainly flouts the distinction Kim explicitly recognizes. The story should go: the putatively higher order *M1* has an instance, $M1_i$, that causes an instance of *M2*, $M2_i$, and does this (according to Kim) by means of an instance ($P1_i$), of its realizing base's causing an instance of *M2*'s realizing base, $P2_i$. Read this way, there is a sense in which we agree with Kim's conclusion: the causal relation between $M1_i$ and $M2_i$ is the same as the causal relation between $P1_i$ and $P2_i$. The picture looks like this:

$M1_i$ -------causes---→ $M2_i$
= is identical with =
$P1_i$ -------causes-----→$P2_i$

Figure B

i.e., causation between mental events just is causation between physical events, since mental events are physical events. But the obvious question now is: why is the supervening property said to be either reducible or causally inert, when the natural assumption, one argued for in preceding sections of this paper, is that the supervening and base properties share instances? If there is just one instance of both the supervening and the base property, then it is true that there is no 'downward causation', where this now means that there are no higher-order *instances* of properties that cause lower-order *instances* of properties. There is no distinction between levels of instances, only between levels of properties. But this is unremarkable, and does not have the consequences drawn by Kim. *This* 'fact' of no downward causation does not lead to the conclusion that the higher-order properties

are causally inert, nor does it lead, *without further argument*, to the conclusion that they are reducible. The causal efficacy of the instance is as secure as the causal efficacy of the base instance, given there is here only one instance. All that is needed to secure the causal power of the supervening property is the plausible additional premise that if a property has instances that are causally efficacious then the property has causal powers. And *if* the higher-order property is irreducible, then it will have independent causal relevance; it will have a causal 'profile' different from that of its particular realizing properties. So what drives Kim to his skeptical conclusion?

Kim's skeptical attitude is to the very idea that properties that are wholly distinct might nevertheless be co-instantiated in a single instance, and it is anchored in his views on the metaphysics of events.[33] These views go beyond commitment to the Property-Exemplification (PEA) Account, since, as we have seen in Section 4, that account can be consistently combined with non-reductive monism. Kim's skeptical attitude is due to his further commitment to (1) the view that mental properties of persons are constitutive properties of the events that are (i.e., are identical with) instances of them, and (2) the view that events have only one constitutive property. Given that physical properties are constitutive properties of the events that are instances of them, the identity condition on events entails that it can only be that $M1_i=P1_i$ if $M1=P1$ (and similarly for $M2_i$ and $P2_i$ and $M2$ and $P2$). Given distinctness of the *M* properties and the *P* ones, distinctness of the instances is assured, and epiphenomenalism, *both* at the level of causal efficacy (of events) *and* at the level of causal relevance, looms.

However, non-reductive monists – even those that commit themselves to the PEA – are free to reject (1) and/or (2) and thereby to block the epiphenomenalist conclusion. If, for example, (2) is rejected, then non-reductive monists can argue that mental/physical events have two constitutive properties, one mental and one physical, both of which need to figure in their identity conditions, and this is possible compatibly with the distinctness of the properties instanced.[34] Taking this line would

[33] Thus, when discussing the relation between the mental property M and it's subvening P, he says "To continue, from *Irreducibility* we have (6) M ≠ P" and notes "…this only means that this instance of M ≠ this instance of P. Does this mean that a Davidsonian "token identity" suffices here? The answer is no: the relevant sense in which an instance of M = an instance of P requires either property identity M = P or some form of reductive relationship between them." (Kim 2005: 42)

[34] One might think this is inconsistent with the existence and identity conditions of events as stated by the PEA account, but it is not (though it is inconsistent with Kim's claims on behalf of that account). As Lombard points out,

leave the relation between mental and physical properties unresolved, however, so we prefer to reject (1): mental properties are not constitutive properties of the events that are instances of them, but rather, supervene on such properties (in the sense specified in Section 4). But this is something any physicalist who thinks that physicalism is true and contingent should do.

There is an argument in Kim (1999) that looks as though it will still deliver the unwelcome conclusion. The critical move is made by the claim that where the realizing relation holds between properties, the instance of the realized property has identical causal powers to that of the instance of the realizer property (so the causal powers of $M1_i$ and $P1_i$ are the same). Kim construes this as flowing from the causal inheritance principle, which says that, in cases of higher-order/lower-order causation, the instance of the higher order-property 'inherits' all its causal power from the instance of the lower-order property. But this causal inheritance principle is not obviously derivable from the less controversial claim that identical instances have identical causal powers, and even this is controversial enough. Let's consider the identity claim first before returning to the inheritance claim.

The identity claim *looks* uncontroversial; indeed, it looks like it provides the ground for the conclusion that the supervening property is causally efficacious, and hence has causal power. It provides support for the efficacy claim because, as we have remarked before, '… is causally efficacious' is an extensional context. If *this* is all that is entailed by the causal inheritance principle, then there can be no objection to it. But there is a way of reading the attribution of causal power to an instance that suggests that it is the *property* instanced, and not the instance itself, whose causal power is in question. What this ambiguity can do is camouflage an inference from the identity of what we will call *instance causal power* to a conclusion about the identity of causal powers of the property instanced. This inference would enable one to move

---

> Suppose that an event, e_1 , is *x*'s exemplifying of *F* at *t*, and that an event, e_2 , is *x*'s exemplifying of *G* at *t*, where *F* and *G* are distinct properties. Despite the fact that Kim's criterion of identity for events says that events are identical only if they are exemplifyings of the same property, that condition does *not* imply that e_1 and e_2 are distinct events. Nothing in that condition or in Kim's existence condition for events says that e_1 could not, in addition to being an exemplifying of *F*, be an exemplifying of *G*, and that e_2 could not, in addition to being an exemplifying of *G*, be an exemplifying of *F*. And if those were the facts, then e_1 and e_2 would be exemplifyings of the same properties by the same objects at the same times, and hence would be, according to Kim's criterion, identical. …that latter idea [that an event can be an exemplifying of only one property] is a consequence, not of the view that events are exemplifyings of properties by objects at times, but of the view that events are explicanda, a

from accepting the picture as presented in Figure B to accepting the picture as presented in Figure A. And it is in fact this further inference that Kim needs in order to arrive at his skeptical conclusion concerning the impossibility of emergent properties. But this inference is infirm, so the skepticism is unwarranted. Additional argument is required in order to be entitled to conclude, from a claim about the identity of the causal power of the instance of co-instanced properties, that the two properties thus co-instanced have the same causal power. In our example, $M1_i = P1_i$, so it is clear that *as instances* they have the same causal power. But that does not by itself license the inference to the conclusion that *M1* and *P1* have the same causal power, since this has to do with instances of *M1* and *P1* in addition to $M1_i$ and $P1_i$. Further, given the possibility of variable realizability, it is clear that we are not entitled to conclude, from the fact that $M1_i$=$P1_i$, that every instance of *M1* is an instance of *P1*.

In the case being considered by Kim it is unlikely that an argument to this conclusion can be mounted that will not beg the question about the coherence of the notion that emergent properties have distinctive causal powers. Ironically, some of the points made by Pettit in favour of program explanation support this view. As we noted earlier, Pettit has made a convincing case for there being supervening properties that are explanatorily *significant*. Such properties are, on the PE model, those that ensure their realization in properties from which one cannot just 'read off' the causal relevance of the higher-order property. The model being examined is one of higher-order property causation, and we are assuming, consistent with Kim's starting point, that the higher-order property is co-instanced in different situations with different base (lower-order) properties. These base properties have, *ex hypothesi*, different causal powers, and, on the plausible assumption that the causal powers of each such property differ from those of any other, it is impossible for the higher-order property to possess the same causal power of each lower-order realizing property. Connected to this is the point alluded to by Pettit, that the difference in causal power can be detected by counterfactualising: a mental property, say, intending to pay that bill, can cause the action of paying the bill, and without that intention the action would not have occurred. But it may be false that without the particular realizer property being instanced, the bill would not have been paid. The intention could have been realized

---

view from which Kim's property-exemplifying account is ultimately derived." (Lombard

by a different base property. So the causal powers of supervening properties can have different profiles from that of the properties supervened upon. There can be emergent properties, properties that are causally relevant to the effects they produce, even though there is no 'pernicious' downward causation.

8. Conclusion.

We have examined Program Explanation and found its metaphysics suspect: resistance to the co-instantiation of realized and realizing properties is unmotivated, and it causes needless problems with the interpretation of the essential 'ensuring' relation. In addition, it renders instances of intentional properties causally inefficacious, a strongly counter-intuitive result. Giving up this non-identity should be palatable to Pettit once the distinction between causal efficacy and causal relevance is recognized and respected, given that his objection to our view appears to depend only on a conflation of the two. We have argued that the preferred alternative re-instates the causal efficacy of the mental while avoiding the problem of causal over-determination. This, however, still leaves open the possibility that, though not causally inert, mental properties may be causally irrelevant, so explanatorily inert. We therefore concluded by addressing an argument of Kim's to the effect that higher-order properties *could not be* causally relevant, as they could not have distinctive causal powers. Kim's conclusion, it was argued, derived from an equivocal premise. Disambiguating that premise destroyed the pessimistic conclusion that such properties are "largely useless for the purpose of causal/explanatory theories" (Kim 1999: 33). There can be higher-order properties with distinctive causal powers, and so non-reductive monism lives on.

Cynthia Macdonald and Graham Macdonald

1986: 55)

References

Brand, M. and Walton, D. (eds.) 1976: *Action Theory*. Dordrecht: D Reidel.

Ehring, D. 1996: "Mental Causation, Determinables and Property Instances", *Nous* 30:4. 461-480

……1999: "Tropeless in Seattle: The Cure For Insomnia", *Analysis* 59, no. 1. 19-24.

Grimes, T. 1991: 'Supervenience, Determination, and Dependency'. *Philosophical Studies* 62:81-92.

Heil, J. 1998: 'Supervenience Deconstructed'. *European Journal of Philosophy* 6: 146-155.

Horgan, T. 1993: 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World'. *Mind* 102: 554-586.

Kim, J. 1976: 'Events as Property Exemplifications'. In M. Brand and D. Walton (eds.) 1976, 159-77.

…….1978: 'Supervenience and Nomological Incommensurables'. *American Philosophical Quarterly* 15: 149-56.

…….1984: 'Concepts of Supervenience'. *Philosophy and Phenomenological Research* 45: 153-76.

…….1993: *Supervenience and Mind*. Cambridge: Cambridge University Press

…….1998: *Mind in a Physical World*. Cambridge, Mass.: MIT Press.

…….1999: 'Making Sense of Emergence'. *Philosophical Studies* 95: 3-36.

…….2003: 'Blocking Causal Drainage and Other Maintenance Chores with Mental Causation'. *Philosophy and Phenomenological Research* 67: 151-176.

…….2005: *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.

Lombard, L. 1986: *Events: a Metaphysical Study*. London: Routledge and Kegan Paul.

Macdonald, C. 1989: *Mind-Body Identity Theories*. London: Routledge.

Macdonald, C. and Macdonald, G. 1986: 'Mental Causation and Explanation of Action'. *Philosophical Quarterly* 36: 145-58.

…….1995: 'How to Be Psychologically Relevant'. In C. Macdonald and G. Macdonald (eds.) 1995, 60-77.

Macdonald, C. and Macdonald, G. (eds.) 1995: *Philosophy of Psychology: Debates on Psychological Explanation*. Oxford: Blackwell Publishers.

Macdonald, C. and Macdonald, G. 2006: 'The Metaphysics of Mental Causation'. *The Journal of Philosophy* 103: 539-576.

Macdonald, G. 1986: 'The Possibility of the Dis-unity of Science'. In G. Macdonald and C. Wright (eds.) 1986, 219-246.

Macdonald, G. and Wright, C. (eds.) 1986: *Fact, Science, and Morality*. Oxford: Basil Blackwell.

McLaughlin, B. 1995: "Varieties of Supervenience", in E. Savellos and U. Yalcin, (eds.) 1995, 16-59.

Melnyk, A. 1995: 'Two Cheers For Reductionism: Or, The Dim Prospects For Non-reductive Materialism'. *Philosophy of Science* 62: 370-388.

Miller, R. 1990: 'Supervenience is a Two-Way-Street'. *The Journal of Philosophy* 87: 695-701.

Pettit, P. 1993: *The Common Mind: An Essay on Psychology, Politics, and Society*. Oxford: Oxford University Press.

Robb, D. 1997: 'The Properties of Mental Causation'. *The Philosophical Quarterly* 47: 178-95.

…….2001: 'Reply to Noordhof on Mental Causation'. *The Philosophical Quarterly* 51: 90-4.

Savellos, E., and Yalcin, U. (eds) 1995: *Supervenience: New Essays*. Cambridge: Cambridge University Press.

Woodward, J. 2003: *Making Things Happen*. Oxford: Oxford University Press.

Yablo, Stephen 1992: 'Mental Causation'. *The Philosophical Review* 101: 245-280.