Emergence in Mind
Cynthia Macdonald and Graham Macdonald

Print publication date: 2010
Print ISBN-13: 9780199583621
Published to Oxford Scholarship Online: May-10
DOI: 10.1093/acprof:oso/9780199583621.001.0001

Emergence and Downward Causation

Cynthia Macdonald

Graham Macdonald

DOI:10.1093/acprof:oso/9780199583621.003.0010

Abstract and Keywords

The principal charge against any emergentist account of the nature of mind is that it leads to incoherence because it is committed to 'downward' causation, and it is this charge that this chapter aims to defeat by appeal to a specific metaphysics of mental causation. Section 1 characterizes some important versions of the doctrine of emergentism. Section 2 develops the challenge anti#emergentists set for advocates of strong emergence, that of demonstrating how emergent properties can be causally effective. Section 3 sets the authors proposal for dealing with the challenge. Section 4 outlines and develops a principled argument against the objection from 'downward causation', dismantling the objection by appealing to the already developed metaphysics. Section 5 defends the authors' position against objections from opponents and argues against the opposing strategy.

Keywords: non#reductive physicalism, property exemplification account, downward causation, higher#level causation, causal relevance, psychological explanation, supervenience, realization

A plethora of recent publications has announced the presence in the world of a large variety of emergent phenomena.[1] According to these, emergence is ubiquitous. According to others, though, it is rather like the Scarlet Pimpernel: we seek it here, we seek it there, we seek it everywhere, but it remains just as elusive. Worse still, there does not seem to be much agreement on what it is that is ubiquitous, or what exactly we are missing when we do not find it.

The fact is that there are a variety of ways in which emergence has been, and can be, conceptualized, and equally, a variety of views on whether it manifests itself in the world (and if so, to what extent). Our aim in this chapter is to outline a particular way of developing an emergentist view of the nature of mind, one that goes under the name of non#reductive monism, and to defend it against some recent objections voiced by those who propose a different way of developing that emergentist view. To this end, we begin in section 1 by characterizing some important versions of the doctrine of emergentism in order to identify some core commitments shared by them and eventually to settle on a version that we will take as 'the' doctrine of emergentism for the purposes of our discussion (a version of what is known as 'strong' emergence). Section 2 develops the challenge anti#emergentists set for advocates of strong emergence and in particular for proponents of non#reductive monism, that of demonstrating how emergent properties can be causally effective. In section 3 we set out our own proposal for dealing with the challenge. Then, in section 4 we outline and develop a principled argument against non#reductive monism based on its emergentist commitments (Kim 2003, 2005), and dismantle the argument by appealing to the metaphysics of our own version of the position. Section 5 defends our position against objections from opponents and argues against the opposing strategy.

The principal charge against the emergentist account of the nature of mind, including non#reductive monism, is that it leads to incoherence because it is (p. 140) committed to 'downward' causation (Kim 1999, 2003, 2005), and it is this charge that we here aim to defeat by appeal to a specific metaphysics of mental causation (Macdonald and Macdonald 2006: Macdonald 2005). Since that metaphysics has come under attack, part of our defence is to address the opposing view. A secondary charge against emergentist accounts is that they cannot explain how emergent properties can have 'new' distinctive causal powers. Though limitations of space makes it impossible for us to address this charge here, elsewhere (Macdonald and Macdonald 1995, 2006) we do so in considerable detail.

1. CONCEPTIONS OF EMERGENCE

Two types of doctrines of emergence have been prevalent in the history of the literature on it. One type, which we shall call 'complex systems emergence', seems to be what many contemporary theorists take to be a doctrine of 'weak emergence'. The other type, which we shall call 'holism emergence,' seems to be what contemporary theorists consider to be a doctrine of 'strong emergence'. Beginning with the distinction between complex systems emergence and holism emergence, we shall work our way toward the distinction between weak and strong emergence.

According to many versions of the doctrine, *complexity* is the key to emergence, the idea being that complex systems exhibit behaviour that is unexpected, or inexplicable, given knowledge only of the simpler parts of the complex whole. Here is one recent characterization of this view:

> complexity emerges at the higher levels of the hierarchy of structure on the basis of the underlying physics, leading to emergent behaviours that cannot be reduced to a description at any lower level. (Ellis 2006: 79)

This usefully incorporates a number of themes relevant to debates about emergence: the notion of hierarchical levels, the thought that underlying the hierarchy is a fundamental physical 'level', and that descriptions of the behaviour of entities at higher levels cannot be *reduced* to descriptions of the behaviour of lower#level entities. Connected to the theme of non#reductibility is *unpredictability* which, in turn, is spoken of in terms of *inexplicability* and causal autonomy, and sometimes this phenomenon of emergence is 'explained' as a result of the *spontaneous* appearance of unprecedented order, such spontaneity itself being said to be the product of the *self#organizing* capacities of the underlying parts.

Given this generous assemblage of characteristics that systems or properties or entities can possess in order to qualify as emergent, it is no surprise to find emergence everywhere, from the behaviour of multicellular slime molds to the operation of financial markets. The price of the coffee one buys in Starbucks (p. 141) will be emergent with respect to the underlying behaviour of the individuals comprising the production and consumption 'base' for coffee.

A second approach to emergence is associated with those who are impressed by what they see as something like a categorical difference between different aspects of reality. This view takes the physical, or the lower level, to be categorized by mechanist, efficient, causality, whereas other levels exhibit some teleological, or final, causality. Here is a quote from one enthusiast, Jan Chistiaan Smuts, in an article on 'holism' in science:

> [Holism] regards natural objects as wholes . . . It looks upon nature as consisting of discrete, concrete bodies and things . . . which are not entirely resolvable into parts; and . . . which are more than the sum of their parts, and the mechanical putting together of their parts will not produce them or account for their characters and behaviour. (Smuts 1929: 640)

Although Smuts says holism views nature as consisting of 'discrete' bodies and events, some contemporary descendants of the view stress the relational nature of the properties mentioned in special scientific explanations, in contrast to the supposedly 'atomistic', or individualistically individuated, entities at the lower level. Sometimes, as in some forms of social holism, the distinction between levels is denied *because*

the putatively 'higher' level is said to be essential for the individuation of the entities at the lower level. So individual people are said to be essentially socially constituted; their psychological characteristics, it is said, cannot be understood except in relation to the social system of which they are a part. (Marx is a famous exemplar of this view: 'human essence is no abstraction inherent in each single individual. In its reality it is the ensemble of social relations' [Marx 1972: 109]).

If one were to represent these two views of emergence as ideal types, one could say the former, complex systems approach stresses the *ordinary workings* of nature as resulting in surprising behaviour of complex systems; the distinction between levels is characterized as a distinction between larger and smaller, or between mereologically composed wholes and their parts, with an emphasis on mundane processes, complicated only because they contain elements of feedback, producing 'resultant' emergence. This contains both a diachronic and synchronic aspect: the later, evolved system contains more than its simpler predecessors, *and* the behaviour of the later system is more complex with respect to its simpler parts and cannot be reduced to the behaviour of those parts.

The second, holistic and relational view is less concerned with what may be called 'unadorned complexity'. In the past, adherents of holism were impressed with the difference that life makes, and so were apt to stress the intricate interconnectedness and interdependence of what may be called 'vital' properties. On this view the difference of levels is *not only* one of larger complexes having properties not reducible to the properties of their simpler parts. Although such irreducibility is endorsed, there is an attempt to make it more substantial, or more (p. 142) principled. In the nineteenth century this produced the doctrine of 'vitalism', a view insisting on the *uniqueness* of the properties that constituted the essence of living forms. The difference of levels was seen as a difference induced not just by the emergence of new properties, but of a new *type* of property. As a consequence some adherents of this view were less inclined to explain this 'emergence' as being the result of simple processes working in unforeseen ways; such an explanation would have been seen as unduly materialist, relying on mechanistic causality which, it was claimed, could never produce the inbuilt *telos* of vital properties and processes. In general this approach emphasized synchronic emergence, the temporal origin of uniqueness being somewhat difficult to explain without invoking a *deus ex machina*.

Perhaps for this reason the holistic view of emergence is not particularly popular today. Even before Darwin and Wallace naturalized biological complexity, and teleology, via natural selection there was a group of scientists in Germany, known as the Munich materialists, who did their best to discredit what they took to be an anti#scientific lobby constituted by Driesch and other vitalists, and did so by insisting that one could

create life in a test tube by mechanical means. Combined with the growing influence of Darwin, the more extreme versions of biological holism gradually withered and died. Nevertheless, we want to keep this view on the table, since it contains, we think, elements of our topic that are crucial to the conception of emergence we want, and think appropriate here, to pursue.

In what follows, we won't have much to say about what we call the 'complexity thesis', mainly because many of the issues here seem to us to be empirical. Our suspicion is that many of the enthusiasts for emergence conflate surprise, unpredictability, and (present) inexplicability with 'in principle' inexplicability. It also seems, to our sceptical eyes, that not even present inexplicability is deemed necessary for this type of emergence to exist. The behaviour of many complex systems is not inexplicable, though it can be surprising. The discovery that the cells composing multicellular slime molds do not have a 'leader', a cell assuming the organizing role, was surprising, given that the behaviour of the multitude of cells seemed to call for such an organizer. But we do now have explanations of how the cells composing slime molds behave the way they do, and why the price of coffee in Starbucks is what it is, and our view is that this type of 'emergent' behaviour does not present any serious concern for the philosopher.

The difference between complex systems emergence and holistic emergence has been characterized by David Chalmers (2006) as a distinction between weak and strong emergence. This distinction has been remarked upon by a number of philosophers and scientists; Herbert Simon, for example, claims that weak emergence is the view that the parts of a complex system have relations that do not exist for the parts in isolation. For example, the template of an enzyme has no function

> until it is placed in an environment of other molecules of a certain kind. Even though the template's function is 'emergent', having no meaning for the isolated enzyme molecule, (p. 143) the binding process, and the forces employed in it, can be given a wholly reductionist explanation in terms of the known physico#chemical properties of the molecules that participate in it. (Simon 1996: 170–1)

Like us, Chalmers thinks that weak emergence contains little of philosophical interest:

> If one is given only the basic rules governing a cellular automaton, then the formation of complex high#level patterns . . . may well be unexpected, so these patterns are weakly emergent. But the formation of these patterns is straightforwardly deducible from the rules (and initial conditions), so these patterns are not strongly emergent. . . . The existence of unexpected phenomena in complex biological systems, for example, does not on its own threaten the completeness of the catalogue of fundamental laws found in physics. (Chalmers 2006: 245)

This leaves us with strong emergence, of which there have again been many characterizations. Chalmers has this to say about it:

> We can say that a high#level phenomenon is *strongly emergent* with respect to a low#level domain . . . when truths concerning that [high# level] phenomenon are not *deducible* even in principle from truths in the low#level domain. (Chalmers 2006: 244)

This is clearly a characterization of a synchronic relation; it is not part of this view that if one cannot deduce truths about the later stages of a process from truths about the earlier stages, then the later stage is strongly emergent. Any output of a non# deterministic process would qualify as strongly emergent if we were to use the above criterion diachronically, and that is clearly not the intended interpretation of those wishing to espouse strong emergence.

Now, we think that this is a very strong version of strong emergence indeed: pack enough into descriptions of the lower domain, including its history, and, arguably, truths about biofunctional properties will be deducible, and so not emergent. It is unsurprising, then, that the only candidate for strong emergence countenanced by Chalmers is consciousness, the notorious explanatory gap ensuring non#deducibility. The rest of the psychological and social world, it is claimed, will be emergent with respect to the physical, but that emergence will be dependent on the physical facts plus consciousness, and so is not *intrinsically* emergent. Given the truths about the physical and the truths about consciousness, the truths of the social, say, will be deducible.

The plausibility of this extreme view will depend to some extent on how it is fleshed out; we need more detail about which truths are included as truths about consciousness in order to assess the claim that given all those truths plus the physical truths, all the rest of the truths about the world are deducible, including truths about the price of Starbucks coffee. We will have no chance of deducing truths about the latter type of fact unless truths about representational facts, about propositional attitudes, are included as truths about consciousness—or so we believe.

(p. 144) Further, there is another famous explanatory gap that is not considered by Chalmers in this context, and that is one constituted by the is–ought gap, or, more generally, the supposed fact–value dichotomy. If there is the notorious Humean gap here, it would appear that there are two options available with respect to moral truths: deny there are any, or admit another strongly emergent domain. One reason for denying there are any is important to our present topic, though it has not been discussed directly in relation to it. This concerns claims about the causal inertness of moral properties (Harman 1977, 1985, 1986); that it is not instances of moral properties, but rather, of our *beliefs* about moral truths, that cause us to behave the way we do, the moral property being at best epiphenomenal. The lack of any causal power is thought to be a

sure sign that there is no *real* moral property doing any causal, and so any explanatory, work.

If this is right, then non#deducibility will need to be supplemented by a more substantial test for emergence, given that emergent properties are claimed to have genuine causal powers. The formulation of strong emergence in terms of the relation between the truths about different domains is what we would call a 'formal' notion of emergence. While the non#deducibility of the higher#level truths is clearly of interest, one would like to know more about *why* the non#deducibility holds. What is it about the *nature* of phenomena and/or properties in the relevant domains that makes for the in principle non#deducibility of the truths of one domain from the truths of the other? Leaving it as a brute fact is unsatisfactory; it invites mystery where there should be none. Our own preference is for a more straightforwardly metaphysical interpretation, one which talks of emergent properties, rather than non#deducible truths, but of course one then needs to say what it is about the properties that makes them emergent. And, again, the problem is that simply asserting their irreducibility to physical properties is unsatisfactory; one needs an argument as to why irreducibility holds.

We think that the argument will be different for different cases (Macdonald 1992; Macdonald and Macdonald 1995). If one wants to defend the irreducibility of biofunctional properties, then one will need to pay attention to what kind of properties these are, note that they have a historical dimension, and that two instances of the same physical#chemical property may be different with respect to whether they are also instances of a specific biofunctional property. If they are thus different, then that physical#chemical property cannot be identified with the biofunctional property with which it sometimes shares an instance, it being a condition on such an identity of properties that they necessarily share all their instances. If one wants to defend the irreducibility of mental properties, then one will need to mount a different argument, one specific to the nature of the mental; and it may need to be different with respect to different types of mental property as well. An argument to the effect (p. 145) that intentional properties are irreducible will be likely to take a different form from an argument to the effect that experiential properties are irreducible. And so on, for other cases, say, that of moral properties.

We do not propose to mount any of these arguments here,[2] since we think that prior work needs to be done: there are powerful arguments purporting to show that irreducibility has *principled* problems, and this is what the present discussion is aimed at tackling. These arguments hark back to our brief discussion of moral properties, and to the thought that, when faced with the non#deducibility of truths about a domain from the truths about a lower#level domain, when faced with an explanatory gap, one has a choice. To put it succinctly, one can be emergentist or eliminativist, be realist or irrealist

with respect to the relevant property#type. Or one can deny irreducibility and non#deducibility. Which way one goes will be case#dependent, but as we have indicated and will shortly discuss in more detail, considerations about causality will be highly relevant.

Given that Chalmers' version of strong emergence is too strong for our taste, and given our preference for a more robustly metaphysical version of the doctrine that focuses on properties rather than on truths, here is our favoured version of strong emergence suitable for the likes of non#reductive monism (i.e., for a physicalist position on the nature of mind):

> A property, *M*, is an emergent property of a (mereologically complex) entity/event, *e*, if and only if:
>
> (1) *M* supervenes on the physical properties, *P*, of *e* (or *e*'s parts).
> (2) *M* is not possessed by any of *e*'s parts.
> (3) *M* is distinct from any structural property of *e*.
> (4) *M* has a causal influence on the behaviour of *e*.[3]

A number of points about these conditions are in order. The first concerns the appeal to supervenience. Emergent properties are said to 'emerge' from other, physical properties of things and, as we have seen, many emergentists believe that the doctrine of emergence is compatible with physicalism. More specifically, non#reductive monism, being committed to the claim that mental properties are distinct from and irreducible to physical ones, claims to be a genuine form of physicalism. The appeal to the notion of supervenience is intended to ward off the charge that mental (and other) properties, being emergent and irreducible, are 'spooky' properties. The idea behind the appeal to supervenience is—to put it in our terms—that even though mental properties are non#physical, they are not worryingly non#physical either; that mental properties, and emergent properties more generally, aren't 'free#floating'. How best to characterize an appropriate relation of supervience for psychophysical and/or other cases is a thorny issue, (p. 146) as is well known: formulations of supervenience abound and we do not propose to enter into an extended discussion of them here.For our purposes, however, supervenience between the mental and the physical can be characterized as 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*. 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**.[4]

A second clarificatory point concerns conditions (2) and (3). These record the fact that an emergent property is one that is not deducible from the properties of any of its parts; it is not 'derivable' from them. Condition (2) makes this clear by explicitly ruling out the possibility that the *M* property of *e* is possessed by any of *e*'s parts; (3) makes it clear by ruling out the possibility that *M* is a structural property. *M* is a structural property of *e* if and only if the proper parts of *e* have properties that are wholly distinct from *M*, and their having those properties is constitutive of *e*'s having *M*.[5] What is at stake in conditions (2) and (3) is the 'distinctive' and 'new' nature of the emergent property.

Condition (4) speaks for itself. An emergent property is one that has distinctive, new, causal powers, powers not possessed by any properties on which it supervenes, nor possessed by any property formed by Boolean operations on such properties. It is this claim that lies at the core of the doctrine of emergence, and which is the source of the charge that emergence leads to incoherence. With this in mind, we turn to one of the most powerful objections to the claim that mental properties are emergent properties.

(p. 147) 2. ANTI#EMERGENCE

The principled argument against strongly emergent properties is one that stresses that emergent properties must come blessed with 'new' causal powers. One rationale for insisting on this lies with 'Alexander's dictum' (Kim 2005), that any real property has distinctive causal powers, so that if emergent properties are to be taken ontologically seriously, they had better have new causal powers.

Again, though, either triviality or falsehood dogs our inquiry. What exactly is required for a new property to have distinctive causal powers? One way of explicating this will be to say that emergent properties can have causal powers, but that these will be essentially dependent on, in the sense of being derivable from, the causal powers of the base properties from which the emergent properties emerge. So, for example, the causal power of a 10kg weight will be different from the causal power of any proper subpart of the weight, but this 'emergent' causal power will be readily derivable from the power of the weight's constituent parts. On this conception, a *causal inheritance* principle is respected: the causal powers of the emerging properties are the product of, or are inherited from, the causal powers of the 'basic' properties. This, the trivially true, version is to be contrasted with the stronger conception of emergent properties as having causal powers whose causal 'action is not detectable at the base level' (Di Francesco 2005) and so not readily derivable from the causal powers of the base properties. This latter view is essentially the one that we have identified as the more appropriate one for our purposes. It has been claimed, plausibly, that any such strong reading requires that the emergent property have powers to influence and control the direction of the lower processes. Such 'downward' causation has been deemed by

Jaegwon Kim to be incoherent and this view of the causal powers of emergent properties false (Kim 1999, 2003, 2005).

Kim's formal characterization of emergence is:

> *Emergence*: Property M is emergent from properties $N_1, \ldots, N_n$ only if (1) M supervenes on $N_1, \ldots, N_n$, and (2) M is not functionally reducible with $N_1, \ldots, N_n$ as its realizers. (Kim 2006: 197)

And he claims that these two clauses capture the concept as it was introduced and intended by the classical emergentists such as Samuel Alexander and C. D. Broad.

> [*Emergence*] can serve as a useful benchmark; any deviation from it is a deviation from the classic conception, and new proposals can be analysed and compared with one another in terms of how far, and in what ways, they deviate from [*emergence*] as a starting point. (Kim 2006: 198)

(p. 148) The difficulty, as Kim sees it, lies with downward causation:

> There is no question that emergentists should want downward causation. Emergent properties must do some serious causal work, and this includes their capacity for projecting causal influence downward. (Kim 2006: 198)

Many advocates of emergence (Alexander 1920; Morgan 1931; Sperry 1980, 1987) require emergent properties to play a significant explanatory role in scientific theory, and epiphenomenal properties, properties that can have no causal influence, cannot play such a role.

As Kim sees it, the deep problem for emergent causal powers arises from the causally/explanatorily closed character of the physical domain, which can be encapsulated as follows:

> *Closure*: If a physical event has a cause, it has a physical cause. And if a physical event has an explanation, it has a physical explanation. (Kim 2006: 199)

The emergentist is thus faced with the challenge: either give compelling reasons for rejecting the closure principle, or demonstrate that downward causally emergent properties are compatible with that principle.

Some emergentists do reject the causal closure of the physical, but we do not intend to go down that route. The challenge, as we see it, is to defend the possibility of emergent causality, consistent with causal/explanatory closure of the physical domain. So our task is to rebut objections to the very possibility that there can be higher#level properties

that have an 'independent' causal profile. In order to do this, we need to spend a bit of time outlining our proposed solution to the possibility of mental causation (Macdonald and Macdonald 1986, 1995, 2006).

The argument for non#reductive monism trades on the distinction between causality, which relates events in extension, and nomologicality, which relates events but only in virtue of certain of their properties. Thus, the first step towards solving the problem of mental causation involves dividing it up and conquering the parts separately. Needless to say, the solutions to the parts had better hang together, and ours not only do so but have the additional advantage of supporting one another. Here, we claim, we have a case of the plausibility#quotient of the whole exceeding the sum of the plausibility#quotients of the separate parts—an example of 'credibility emergentism'.

As we see it, 'the' problem of mental causation consists of three parts. The first concerns the causal efficacy of mental events, the second concerns what we have called the causal relevance of mental properties (the so#called *qua problem*), and the third deals with the compatibility of different levels of causation, the problem of downward causation.

As noted, the problem of efficacy concerns causality taken in extension; two events can be causally related even though the way in which that causal relation (p. 149) is described is explanatorily unilluminating. One of the clearest examples of this is provided by the description of the event that caused *b* as 'the cause of *b*'; such a description in a true causal statement will pick out a property of the cause, but will be explanatorily useless. Nevertheless, the event so described will indeed be the cause of *b*, and so will be causally efficacious in bringing about the *b* event.

Clearly, then, there can be true causal claims that yield nothing useful in the way of explanation. The second component of the 'problem', however—the problem of causal relevance of mental properties—does concern explanation; the challenge is to show that event *a* is causally relevant to event *b* by specifying *a* in a way that can explain why *b* occurred. In the language that has become mandatory, the solution to the second part of the problem is to show how *a* causes *b qua a*'s possessing property *M*. As has become clear, not just any property specification will do the required explanatory work, and it is a matter of ongoing debate how best to locate those properties in virtue of which an effect occurs. For present purposes we simply need to point out that successfully doing this is different from, because it involves more than, solving the difficulty concerning causal efficacy. The conflation of these two problems has been responsible for much of the confusion surrounding the problem of mental causal efficacy, or so we believe (cf. Yablo 1992; for more on causal relevance see Macdonald and Macdonald 2006).

The third part of any satisfactory account of mental causation is the one we have seen as especially relevant to any hope of defending a substantial doctrine of emergentism, the

possibility of downward causation. The crucial 'downward causation' claim here is that mental events can cause physical effects. If mental properties are understood as being higher#level ones, then this can make it appear that that claim requires that there be downward causation.

In what follows we indicate first, as briefly as possible, how our solution overcomes problems of causal efficacy; we then move on to deal with downward causation, as that is most relevant in answering the critics of emergentism. We conclude by considering objections to our solution based on an opposing emergentist view of mental causation.

3. CAUSAL CONSIDERATIONS

Our solution to the problem of causal efficacy of mental events pays special attention to the distinction between properties, construed as abstract and universal, and their instancings or exemplifyings, understood as events, where such instancings are not to be understood as tropes. It appeals to a version of the property exemplification account (PEA), *developed in a particular way*. According to this, events, such as Jones's running now, not only have properties, such as the property of being a running event, but are the exemplifyings of properties, (p. 150) such as the property *runs*. That is to say, they are identical with exemplifyings of (*n*#adic) act#or event properties at (or during intervals of) times in objects.[6] 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. In our example, the event of Jones's running now just is the exemplifying in Jones of a property of Jones, the property, *runs*, now. Such properties are sometimes termed constitutive properties of events, and are so termed because they are the properties of subjects whose exemplifyings in those subjects just are events. So when it is said that events 'have' constitutive properties, this is not to be understood as the claim that they possess such properties.[7]

Events construed along these lines are sometimes referred to as 'structured particulars', and are so deemed because they have not only constitutive properties, but also constitutive objects (or subjects) and constitutive times. That is to say, (p. 151) 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:[8]

> *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*',

where expressions of the form '[x,P,t]' are construed as singular terms referring to events.[9] Kim takes expressions of this form to be canonical descriptions of events because they pick such events out in terms of their constitutive objects, properties, and times. Given that events can be described in ways that do not pick them out in terms of their constitutive 'components', there is no easy way of telling, for any given description of an event, whether it is a canonical description of that event. The importance of this point will emerge in our discussion to follow.

Although Kim himself assumes that events have only one, unique, constitutive property, the view that events are property exemplifications does not require this commitment, nor do the existence and the identity conditions, as formulated above, since neither condition states that events whose constitutive (p. 152) properties are specified by singular terms referring to them do not possess *other* constitutive properties in addition to the ones specified by those terms. If they do, then such events will have more than one canonical description, and the identity condition will require that such events have *all* the same canonical descriptions.[10] Indeed, the version of the account we favour, developed by Lawrence Lombard (1986), explicitly allows for an event's having more than one constitutive property. This version is predicated on the assumption that events are paradigmatically and fundamentally changes, where these are *not* to be understood as states or persisting conditions. This assumption is founded on the intuition that some properties are such that their possession by an object at a time implies change, whereas others are not. Lombard labels these two sorts of properties 'dynamic' and 'static' respectively, and argues that only exemplifyings of the dynamic ones imply the existence of events.

If a material substance has a dynamic property during an interval of time, then it will be true that that substance is changing during that interval from having one static property to having another. This will be true because a dynamic property just *is* the property of first having one, then another, static property. Thus Lombard's version of the PEA, unlike Kim's, not only countenances the possibility that an event may have more than one constitutive property, but actually requires it.

As the above discussion suggests, in addition to constitutive properties, events also have characterizing properties. These are properties that events themselves possess, at least some of which they possess in virtue of having the constitutive properties they have. Thus, for example, the event that is the exemplifying of the property, *runs*, in Jones at time *t*, has as its constitutive property a property of Jones. That event has the property of being a running.

The PEA construes properties as abstract, multiply#exemplifiable entities that can have, but are not identical with, their exemplifyings. According to it, to say that mental events are identical with physical events 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 (p. 153) a physical property of that subject in that subject at that time. Crucially, this amounts to the claim that there is just *one exemplifying* of two properties, one mental, and one physical, by an object at a time.[11] That this is possible is apparent from determinable/determinate examples, such as that of being coloured and being red. The most natural understanding of the relation between these properties is that for an object to instance the latter (being red) just is for it to instance the former (being coloured): nothing further is required, once the latter is instanced, for the former to be instanced. Unlike the determinable/determinate property relation, the relation between mental and physical properties is not both metaphysical and conceptual. However, if non#reductive monism is committed to the view that mental properties supervene on physical ones in the sense specified in section 2, the result is that mental properties of persons *are not themselves constitutive properties of the events that are (identical with) exemplifyings of them*, but rather, supervene on those events' physical, constitutive properties. That is to say, a description such as 'Jones's having pain at *t*' (i.e., '[*Jones*, *having pain*, *t*]') is *not* a canonical description of the event which is Jones's having pain at *t*. And our view is that although the supervenience relation is a weaker metaphysical relation than the determinable/determinate one, both are cases where there can be a single exemplifying of distinct properties.

Thus, appealing to the PEA in order to rescue causal *efficacy* for mental events requires simply recognizing that an event can be a single exemplifying of both a mental property and a physical property. In the case of mental and physical properties of events, we claim that this is just what happens, and hence, that the following 'Co#Instantiation Thesis' for events is true:

> (CI) Two or more properties of an event can be co#instantiated in a single instance, i.e., there can be just one instance of distinct properties. (Macdonald and Macdonald 1986, 1995)

By the extensionality of the causal relation, if the physical event is causally efficacious, the mental one is. This shows that the PEA has the resources with (p. 154) which to rescue the causal efficacy of instances of mental properties, which on any plausible account is necessary for the causal relevance of the properties themselves.

4. DOWNWARD CAUSATION*

The argument against emergent properties was tied to the incoherence of downward causation. We have accepted that some account of downward causation must be given, and it must respect the causal closure of the physical. 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) level. These effects will be higher#level effects (given that emergent properties themselves are higher#level). But, given that supervenience holds, this means that the higher#level effects will have lower#level realizations. So, it is claimed, it is by causing instances of the lower#level (base) realizing properties that an emergent property will cause a higher#level effect. So, higher#level causation presupposes 'downward' causation.

Why, according to Kim, is downwards causation incoherent? Consider emergent properties *M1* and *M2*, where *M1*'s instantiation 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, and this is the 'Downward Causation' conclusion. This conclusion holds for all higher#level supposedly autonomous causal action: given supervenience, 'we can no longer isolate causal relations within levels; any causal relation at level *L* (higher than the bottom level) entails a cross#level, *L* to *L#1*, causal relation' (Kim 2005: 40).

But given that *M1* is realized by *P1*, and given irreducibility (i.e., that *M1≠P1*), we now have two sufficient causes of *P2*, and this breaches the spirit of the principle of closure, which allows only one sufficient cause. Given physicalism, we are driven to the conclusion endorsed by Kim: 'The putative mental cause, *M1*, is excluded by the physical cause, *P1*. That is, *P1*, not *M1*, is the cause of *P2*'[12] (Kim 2005: 43). It follows that the emergent property *M1* is not independently causing *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.

(p. 155) It looks as though Kim has a sound argument for the causal irrelevance of emergent properties. For the emergentist, only triviality looms. But we claim that the argument is not sound. In this section we want to show that there is a sense in which it is true that downwards 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.

What is of particular interest in the way the argument is set up is that it shuttles between talk of the downward causal power of properties and that of their instances. It is not that Kim is unaware of the importance of the instance–property 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)

With this distinction in mind, we can express what is happening in putative 'downward causation' as: an instance of the lower#level property *P2* is caused by an instance of the higher#level *M1*, and *M1* does this while being realized by *P1*. 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.[13]

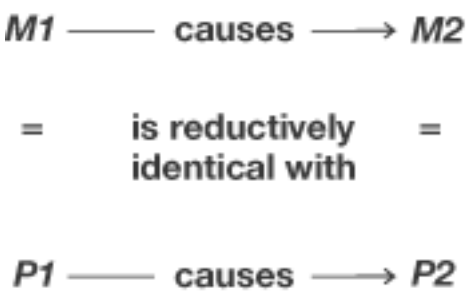

Figure 10.1.

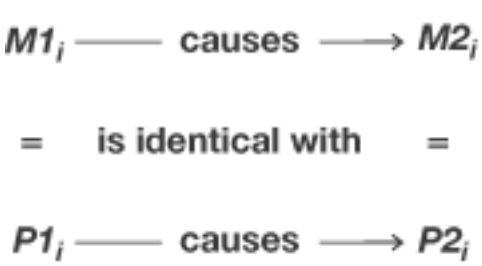

Figure 10.2.

Diagramically, his picture of the situation is this (Kim 2005: 55): 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 while being identical with an instance of its realizing base, $P1_i$, thus 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: (p. 156) That is, causation between mental events just *is* causation between physical events, since, given physicalism, 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 earlier, is that the supervening and base properties are instantiated in a single instance? If there is just one instance of both the supervening and the base property, then it is true that there is no 'downwards causation', where this now means that there are *no higher#level instances* of properties that have as effects

lower#level instances of properties. This, though, is because there is *no distinction of levels of instances*, only levels of properties; at the level of instances, the world is flat. But this is unremarkable, and does not have the consequences drawn by Kim. This 'fact' of no downwards causation does not by itself lead to the conclusion that the higher#level 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#level property is irreducible, then it will have independent causal relevance; it will have a causal 'profile' different from that of its particular realizing properties.

It may be objected that the higher#level mental property *is* reducible, but this has only been asserted as a way of rescuing its causal efficacy and relevance. Since, however, we have defused that argument, there is nothing to stand in the way of the irreducibility claim, and familiar points about the multiple realization of mental properties support irreducibility. So the causal powers of supervening properties can have different profiles from those of the base properties on which they supervene. There can be emergent properties, properties that are causally relevant to the effects they produce, even though there is no 'pernicious' downwards causation.

What, then, drives Kim to his sceptical conclusion? There is an argument in Kim (1998) that looks as though it will still deliver the unwelcome conclusion.The critical move is made by the claim that where the realization relation holds between properties, the instance of the realized property has identical causal powers to that of the instance of the realizer property, so that, in a situation in which *P1* realizes *M1*, the causal powers of $M1_i$ and $P1_i$ will be the same. Now, Kim (p. 157) 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.[14] 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. On our account, this reading of the causal inheritance principle is clearly unexceptionable, given that

the most plausible view is that the realized and realizing properties have the same instance. 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 from accepting the picture as presented in Figure 10.2 to accepting the picture as presented in Figure 10.1. And it is in fact this further inference that Kim needs in order to arrive at his sceptical conclusion concerning the incoherence of downward causation, and hence the rejection of emergent properties. But this inference is infirm, so the scepticism 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 instanced have the same causal power. The identity of instance causal power in our example ($CM1_i = CP1_i$) does not by itself license the inference to the conclusion that *M1* and *P1* have the same causal power, since this latter causal power, property causal power, is connected with instances of *M1* and *P1* other (p. 158) than $M1_i$ and $P1_i$. Further, given the possibility of multiple realization, 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 simply beg the question about the coherence of the notion that emergent properties have distinctive causal powers.

5. OBJECTIONS FROM THE OPPOSITION

We believe that our solution to the problem of downward causation is both simple and elegant, offers a clear and coherent account of how higher#level properties can be causally effective without overdetermining the effects they and their realizer properties have in the world, and makes much needed sense of what the relation between higher#level and lower#level properties is when these are related as realized to realizer. However, objections have come forward from sources that take a different emergentist view of the relation that mental and other higher#level properties bear to their physical realizers and that claim to make better sense of how emergent properties can exert a 'downward' causal influence on the workings of the physical domain. One very well known source of this view is to be found in the writings on 'program explanation' of Philip Pettit (1993, 2007; Jackson and Pettit 1990); another, strikingly similar view has been voiced in a number of writings by Carl Gillett (2002, 2006a, 2006b).[15] A key theme in both is that emergent properties exert their causal influence, not by being co#instantiated with their physical realizer properties, but by *non#causally ensuring* that the realizer properties will be instantiated, and, by so ensuring, themselves will be

causally 'effective'. In Pettit's terminology, the realized properties, by ensuring that their realizers will be instantiated, 'program' for certain effects that instances of those realizer properties would not, had they not stood in this relation to the realized ones, have had. In Gillet's terminology, the realized property instance non#causally determines some of the causal powers that its realizer property instance contributes to individuals that have it. Such powers are 'conditioned' powers, ones conditioned by the presence, and instantiation, of the emergent, realized properties.[16]

(p. 159) Given, as we have seen, the importance of the property/instance distinction in discussions of the causal efficacy and 'downward causation' of mental and other higher#level properties, the question of prime importance here is how it is that the realized properties can non#causally ensure or determine that the realizer property will be instantiated, thereby bringing about an effect that otherwise it would not have brought about. Both Pettit and Gillett acknowledge the property/instance distinction (although Gillett in particular is not careful to observe it[17]), as well as the point that properties, being abstract entities, are not the sort of things that can be causes. Their view is that it is instances of properties that are causally efficacious. So, in the case of realizers of mental (and other higher#level) properties, it is their instances that are causally efficacious. The question is how it is that an instance of a higher#level property can 'non#causally ensure/determine' that an instance of the realizer property will have an effect that it would otherwise not have brought about without being epiphenomenal.

In his earlier work, Pettit does not tell us how it can do so, except to say that it does not do so by being co#instantiated with an instance of a realizer property, as our solution would have it. Indeed, he objects to our position precisely on the grounds of its commitment to the co#instantiation principle.[18] Fundamentally, (p. 160) the Program Explanation (or 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#level 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#level (mental) property 'ensures that' a lower#level (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#level 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)[19]

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 overdetermination.

Like Pettit, Gillet doesn't really give us much of an idea as to how instances of emergent properties non#causally determine or ensure that their 'realizers'—instances of their realizer properties—will occur without being epiphenomenal. He does say that the relation between the emergent property instance and its realizer property instances is a 'part#whole' one, this being so because the former is identical with a combination, or sum, of instances of properties and relations between them that realize the emergent one (Gillett 2006b). Since the aggregate property instance just is a sum or collection of instances of the microphysical properties and relations between them that realize the emergent one, and the emergent property instance is identical with the aggregate property instance, the emergent property instance bears a part#whole relation to the instances of the properties that constitute the aggregate property instance. However, the mere fact that the instances of microphysical properties (p. 161) that constitute the realizing property aggregate instance bear a part#whole relation to the emergent property instance does not by itself explain how the latter is, or can be, causally effective, since the part#whole relation is not a causal relation. Further, since the emergent property instance is identical with the combination of instances of microphysical properties that realize the emergent one, it is hard to see how any genuinely new causal powers are contributed by the emergent one to individuals that instance the microphysical property#realizers.

So neither version of the alternative emergentist doctrine gives a genuine alternative account of how 'downward' causation is possible, or occurs.

We think that these proposed alternatives to the co#instantiation model are fuelled by two thoughts. One is that, given that emergent properties are distinct from and irreducible to their realizer properties (where 'properties' means 'property types', not 'property instances'), their instances must also be distinct. The second is that, given the distinctness of the instances of such properties, one can only avoid troublesome overdetermination problems, and with it the threat of epiphenomenalism to the emergent ones, by denying that the 'influence' that they exert is genuinely causal. The consequence is a peculiar mixture: emergent property instances are indeed 'causally effective' but they are so, not by actually causing their realizer instances, but rather, by non#causally ensuring that those instances occur. We think that, in effect, this is to concede epiphenomenalism, since the emergent property instances are not themselves

causally efficacious. This being so, it is hardly a solution to the problem of mental causation, nor, more generally, to the problem of 'downward' causation of emergent properties.

There is a suggestion in recent work by Pettit (2007) that the program model can accept the co#instantiation in some cases (the ones that we bring to bear on the discussion of the problem of mental causation) but has the virtue of being more general than our solution in allowing for programming to occur in cases of causation involving realized and realizer properties where, it is claimed, the co#instantiation model will not work, via distinct instances of the higher#level and realizer properties. Though we do not see how it can accommodate the co#instantiation without robbing it of the distinctive appeal over our solution claimed for it, what is important is the claim that the program solution handles other cases with ease that cannot be handled by the co#instantiation one. For it is here where we and Pettit disagree about the virtues of the program model and about whether it can do distinctive work in the area of higher#level causation.

We claim that our solution is by far the most plausible and does the work required of it in accounting for the causal efficacy of higher#level, emergent properties, in cases that involve properties in the domains of the special sciences, where those properties are best understood as standing as realized to physical realizers: biofunctional properties in biology, intentional properties in psychology, and so on. Pettit, however, thinks that there are cases of ordinary macrophysical causation, where macrophysical properties are also best understood as standing (p. 162) in a relation of realized to physical microphysical property realizers, but where the co#instantiation model does not get the metaphysics of the situation right. Now, we have not maintained that this model generalizes to all cases of higher#level/lower#level 'causation'; it may be that it will work only for cases of higher#level properties in the special sciences (and that might be good enough for us!). Nevertheless, in closing, we'd like to say something in response to the particular example Pettit brings to bear on the debate between us, since we don't think that it establishes what he thinks it does.

Like us, Pettit thinks that there are two 'problems' of mental causation; one concerns causal efficacy of property instances, and the other concerns the causal relevance of properties themselves (the *qua problem*). And, like us, he thinks causal relevance concerns in part causal efficacy, in part explanatory potential. Finally, like us, he thinks that causal relevance concerns the potential of the mental (or other higher#level property) to explain, not why a particular effect, under just any description, occurred, but why an effect of a certain *type* occurred. His claim is that the program model is designed to handle the problem of causal relevance, not causal efficacy.

Where he parts company with us is in his understanding of the sense in which causal relevance concerns in part causal efficacy. For us, the property that is causally relevant

must itself have instances that are themselves causally efficacious; that is an absolute requirement on causal relevance. When Pettit supposes that he can accept this within the confines of the program model, then, he is mistaken, *for the program model does not require this*. Pettit makes this clear in both his recent and his earlier work. And it is this that is the source of dispute between us. Bearing it in mind, let's look at the example that Pettit thinks will not fit the co#instantiation model, see what our model can say about it, and compare it with what Pettit's own model can say about it.

Suppose that there is a closed glass flask containing water that is boiling, with a mean molecular motion of such and such. And suppose that one of the molecules in this, in the aggregate of molecules that has the property of boiling, is moving with a particular momentum and position in such a way that it breaks a molecular bond in the flask. As a result of this, the flask cracks.

Pettit says that, intuitively speaking, the property of being at boiling temperature programmes for the 'production of the breaking' and that this is a property of an aggregate of molecules, though the production of that effect—the cracking of the flask—is not brought about by an event involving the entire aggregate but only by an event involving one of the molecules in the aggregate, one of its components. This programming property might be realized in any number of ways, by different numbers of molecules having different momentum#position properties, but in each case there will be one molecule in any such aggregate whose instancing of a particular momentum#position is sufficient to (p. 163) break the bond in the glass and thus to crack the flask.[20] In this situation, Pettit says,

> A property of the component event—the momentum#position property—programs at the same time, but in a more specific manner, for the production of that very same effect. And the more general programmer programs for the effect via the programming of the more specific programmer. The program architecture still holds. . . . It is important to see that the program model may be extended to more complex cases like this . . . in many of these cases it would require procrustean efforts of reconstrual to be able to argue that all the relevant programming properties are co#instantiated. (2006: 223)

Pettit's basic complaint here seems to be that, because the relation between water and the constituent molecules of the aggregate with which (he seems also to hold) water is identical is a part#whole, or mereological, one, and because only one of these constituent molecules instantiates the property that is causally relevant to the cracking of the flask and whose instance is causally efficacious in bringing that effect about, although the entire aggregate has the property of being at boiling temperature, the instance of being at boiling temperature (instantiated by the aggregate as a whole)

cannot be construed as identical with the instance of having a certain momentum#position, a property had by the constituent molecule. Instead, the instance of the property, being at boiling temperature (the programming property), non#causally ensures that the property of having a certain momentum#position will be instantiated, thereby bringing about the cracking of the flask.

This still leaves unresolved how it is that an instance of the programming property non#causally ensures that the realizing property will be instantiated. Given that Pettit rejects the co#instantiation solution, his explanation is that

> The event that realizes the more general programming property involves the mass of water molecules, and the event that realizes the more specific programming property involves a part of that whole: the particular molecule that does the damage. The instance of the property of the whole—the boiling of the mass of molecules—will be one event, the instance of the property of the part—the vibration of the efficacious molecule—will be another; they will be non#identical. But they will still not be distinct events. This appears in the fact that the change in the whole cannot cause the change in the part; the change in the whole is partly constituted by the change in the part: it is superveniently determined by the changes of motion in that and other parts. (2006: 224)

What do we have to say about this example? In contrast to Pettit, we think that it is not at all clear that there is no case of co#instantiation here, and in other cases like it. In order to establish this, Pettit would need to show that in general, (p. 164) when an object with parts changes in such a way as to bring about an effect of a certain type, and it changes solely because one of its parts changes, in order for the change in the whole to be identical with the change in the part, the change in the part must be of the same type as the change in the whole. But this seems false. Suppose, for example, that I stammer, and that it is because my stammering is a stammering, i.e., is an instance of the property of being a stammering, that my conversational partner blushes with embarrassment. But my stammering just is—is identical with—my tongue's catching, so that the property of being a stammering and that of being a tongue#catching are here co#instantiated.

It is true that being a tongue#catching will not causally explain why my partner blushed with embarrassment. In our terminology, that property is not causally relevant to that effect, as described. We think that it is this kind of consideration that leads Pettit to say that the property of the whole (here the property of being a stammering; in his own example, the property of being at boiling temperature) programs for but is not co#instantiated with the property of one of its parts (the tongue#catching, and the property of being a tongue#catching; in his example, the constituent molecule's vibrating, and the

property of being a vibrating) that is effective in bringing about the effect. But this seems to us just to confuse causal efficacy with causal relevance. Note that our example, like his, concentrates on an ordinary case of macrocausation.

So our response is to issue a challenge: Pettit's claim that the co#instantiation model won't work in a case such as his trades on an assumption that seems to us to be false, and our stammering/tongue#catching example brings this out. More generally, however, we are not convinced that Pettit's example serves his own model of an emergent property exerting a 'downward' influence, and so we are not convinced that this is a case where the problem of emergent causation arises. The issue depends on what role the property of having a mean molecular motion is playing in the example and argument.

As we indicated, Pettit's assumption seems to be that water is identical with an aggregate of molecules, and, along with this, that the event that is water's boiling is identical with 'an aggregate event', the momentum#positions of the molecules. The instance of the property of being at boiling temperature, the property which he says is a property of the aggregate of molecules, seems to be either identical with, or realized by, the instance of the property of having a mean molecular motion. The most plausible understanding of the example, as he describes it, is that the instance of being at boiling temperature is identical with the instance of having a mean molecular motion, especially if the boiling is identical with the aggregate event.

Given this understanding of the relation between the two property instances, the claim is that being at boiling temperature/having mean molecular motion, programs for the production of the cracking of the flask by virtue of its instance (= the instance of the property of having mean molecular motion) non#causally ensuring that an instance of one of its realizer properties, the property (had by (p. 165) one of the molecules constituting the aggregate constituent molecules) of having a certain momentum# position, will break a particular bond in the surface of the glass, thereby cracking it. The non#causal ensuring is accounted for by the fact that the realizing instance bears a part#whole relation to the realizer, higher#level instance.

It seems to us that this explanation, like Gillett's, effectively concedes that the higher# level macrophysical property is epiphenomenal, since the part#whole relation is not a causal one. Further, since the higher#level macrophysical property instance is identical with the combination of instances of microphysical properties that realize it, it is hard to see how any genuinely new causal powers are contributed by the former. In contrast with the view that the higher#level macrophysical property is emergent, it seems rather to be a good example of a resultant property, one whose causal powers are a product of, or function of, the causal powers of its realizer properties, and so a case of 'upward' rather than 'downward' causation.

A resultant property of a complex entity, Kim tells us, is arrived at by 'mere' addition or subtraction of properties of its parts or components (Kim 2006).Examples of such properties are those of the shape of a table, its mass, or its weight; properties that are 'resultants' of the properties of the table's microstructure. This notion of being 'additive' covers different kinds of case. One is where the property of the whole is a function of properties of the *same* type had by the constituents. An example is Pettit's case of the property of having mean molecular energy and having a certain momentum position. Another kind of case is one where, although the 'resultant' property of the whole is a function of the properties of its microstructure, the properties of the microstructure in virtue of which the object has its resultant property are not properties of the same type. So, for example, this 50kg table's having the property of weighing 50kg is a result of the properties of its microstructure; the latter properties are not themselves weight properties.

In both kinds of case it seems that the nature of the resultant property is fixed entirely by the properties out of which it is constituted, and the entity's having it doesn't introduce a pattern of behaviour at the micro# (or lower#) level that differs from the sort of behaviour that would occur in its absence. This being so, the property is said not to be emergent in the sense that it contributes new causal powers that require new laws or forces not already in place at the lower, realizing level.

Pettit's example is one, not only of a resultant property, but of an 'averaging' property. This makes it unsuitable for the purposes of adjudicating between different emergentist versions of non#reductive monism, one based on the program model, the other based on the co#instantiation model. The reason is that an averaging property, even more obviously than other resultant properties, is one that a complex entity has solely in virtue of the properties of its constituents, and one whose causal powers are exhausted by those of its constituent properties. (p. 166) It is therefore not the type of property that raises problems for emergentism generally or for non#reductive monism specifically. It could not raise the threat of overdetermination, the very threat that lies at the core of emergentist views. That threat is specifically one concerning the coherence of 'downward' causation.

REFERENCES

Alexander, S. 1920. *Space, Time, and Deity*, 2 vols. London: Macmillan.

Armstrong, D. 1978. *Universals and Scientific Realism*. Vol. 2: *A Theory of Universals*. Cambridge: Cambridge University Press.

Bigelow, J. and Pargetter, R. 1989. 'A Theory of Structural Universals'. *Australasian Journal of Philosophy* 67: 1–11.

Chalmers, D. 2006. 'Strong and Weak Emergence'. In P. Clayton and P. Davies (eds), *The Re#Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*. Oxford: Oxford University Press, 244–54.

Clayton, P. and Davies, P. (eds) 2006. *The Re#Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*. Oxford: Oxford University Press.

Di Francesco, M. 2005. 'Filling the Gap, or Jumping Over it? Emergentism and Naturalism'. *Epistemologia* 28: 95–122.

Ellis, G. 2006. 'On the Nature of Emergent Reality'. In P. Clayton and P. Davies (eds), *The Re#Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*. Oxford: Oxford University Press, 79–107.

Gillett, C. 2002. 'Strong Emergence as a Defense of Non#Reductive Physicalism: A Physicalist Metaphysics for "Downward' Determination" '. *Principia* 6: 89–120.

—— 2006a. 'Samuel Alexander's Emergentism: Or, Higher Causation for Physicalists'. *Synthese* 153: 261–96.

—— 2006b. 'The Hidden Battles Over Emergence'. In P. Clayton and Z. Simpson (eds), *Oxford Handbook of Religion and Science*. Oxford: Oxford University Press, 261–96.

Harman, G. 1977. 'Ethics and Observation', in G. Harman, *The Nature of Morality*. New York: Oxford University Press, 1–10.

Reprinted in G. Sayre#McCord (ed.), *Essays on Moral Realism*. Ithaca, New York: Cornell University Press, 1988, pp. 119–24.

—— 1985. 'Is There a Single True Morality?'. In D. Copp and D. Zimmerman (eds), *Morality, Reason and Truth*. Totowa, NJ: Rowan and Allanheld, 27–48.

—— 1986. 'Moral Explanations of Natural Facts: Can Moral Claims be Tested?'. In N. Gillespie (ed.). *Southern Journal of Philosophy* 24 (Supplement on Spindel Conference, 1986: *Moral Realism*): 57–68.

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

Johnson, S. 2002. *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. New York: Touchstone.

Kim, J. 1972. 'Phenomenal Properties, Psychophysical Laws, and the Identity Theory'. *Monist* 56: 177–92.

—— 1973. 'Causation, Nomic Subsumption, and the Concept of Event'. *Journal of Philosophy* 70: 217–36.

(p. 167) —— 1976. 'Events as Property Exemplifications'. In M. Brand and D. Walton (eds), *Action Theory*. Dordrecht: D. Reidel, 159–77.

—— '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, MA: 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: 151–76.

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

—— 2006. 'Being Realistic About Emergence'. In P. Clayton and P. Davies (eds), *The Re#Emergence of Emergence: the Emergentist Hypothesis from Science to Religion*. Oxford: Oxford University Press, 189–202.

Lennon, K. and D. Charles. 1992. *Reduction, Explanation, and Realism*. New York: Oxford University Press.

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

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

—— 2005. *Varieties of Things: Foundations of Contemporary Metaphysics*. Oxford: Basil Blackwell.

—— 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), *Philosophy of Psychology: Debates on Psychological Explanation*. Oxford: Basil Blackwell, 60–77.

—— 2006. 'The Metaphysics of Mental Causation'. *Journal of Philosophy* 103: 539–76.

—— 2007. 'Beyond Program Explanation' in G. Brennan, R. Goodin, F. Jackson, and M. Smith (eds), *Common Minds: Essays in Honour of Philip Pettit*. Oxford: Oxford University Press, 1–27.

Macdonald, G. 1986. 'The Possibility of the Dis#unity of Science'. In G. Macdonald and C. Wright (eds), *Fact, Science, and Morality*. Oxford: Basil Blackwell, 219–46.

—— 1992. 'Reduction and Evolutionary Biology'. In K. Lennon and D. Charles (eds), *Reduction, Explanation, and Realism*. New York: Oxford University Press, 69–96.

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

Marx, K. 1972. 'Theses on Feuerbach'. In R. Tucker (ed.), *The Marx#Engels Reader*. New York: W.W.W. Norton & Co, 107–9.

McLaughlin, B. 1995. 'Varieties of Supervenience'. In E. Savellos and U. Yalcin (eds), *Supervenience: New Essays*. Cambridge: Cambridge University Press, 16–59.

Morgan, C. Lloyd. 1931. *Emergent Evolution*. New York: Henry Holt.

Morowitz, H. 2002. *The Emergence of Everything: How the World Became Complex*. New York: Oxford University Press.

O'Connor, T. 1994. 'Emergent Properties'. *American Philosophical Quarterly* 31: 91–104.

—— and Wong, Hong Yu. 2005. 'The Metaphysics of Emergence'. *Noûs*, 39: 658–78.

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

(p. 168) Pettit, P. 2007. 'Joining the Dots'. In G. Brennan, R. Goodin, and M. Smith (eds), *Common Minds: Essays in Honour of Philip Pettit*. Oxford: Oxford University Press, 215–338.

Simon, H. A. 1996. *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Smuts, J. C. 1929. 'Holism'. *Encyclopedia Britannica*, 14th edn, vol. 11: 640.

Sperry, R. 1980. 'Mind#Brain Interaction: Mentalism, Yes; Dualism, No'. *Neuroscience* 5: 195–206.

—— 1987. 'Consciousness and Causality'. In R. Gregory (ed.), *The Oxford Companion to the Mind*. Oxford: Oxford University Press, 164–66.

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

Notes:

(*) This work has been supported by a grant from the Royal Society of New Zealand Marsden Fund.

(1) See, for example, Johnson 2002; Morowitz 2002; Clayton and Davies 2006.

(2) For discussions relevant to the mental case see Macdonald (1992) and Macdonald and Macdonald (1995, 2006).

(3) This characterization of an emergent property is an adaptation of one advanced by O'Connor (1994).

(4) This is an adaptation of the definition of strong supervenience given by 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 his, though his 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 variable realizability of mental properties.

(5) This is a version of O'Connor's (1994: 5) formulation, adapted from Armstrong (1978). The formulation allows for the possibility that a simple conjunctive property A&B of emergent properties A and B might be non#structural (though non#basic). Although O'Connor here speaks of the having by e's parts of certain properties being constitutive of e's having the structural property M, other comments in this work and in more recent work (O'Connor and Wong 2005) clearly indicate that by 'constitutive of' he means 'ontologically reducible to': "there is *nothing more* to having the structural property than being composed by parts having certain relations to one another—it is ontologically reducible" (2005: 10). This contrasts with Bigelow and Pargetter (1989),

whose theory of structural properties takes 'constitutive of' to mean 'essential to but ontologically distinct from'.

(6) In the terminology preferred by Kim, whose version of the account we describe and develop further here, events are *exemplifications* of properties by objects at times (see Kim 1976). But Kim himself, and many others who take a universalist rather than a trope view of properties, often use the term 'instance' as an alternative 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' (for reasons akin to those given by Lawrence Lombard [1986]), since it makes clear that events are fundamentally changes, whose 'constitutive' properties are dynamic rather than static, or its cognate term, 'instancings', since we think that failure to do so blurs the crucial distinction between a substance and an event (see Macdonald 1989, ch. 4)).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 entity, 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. 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, 2005).

We now 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 (e.g. Pettit [1993] and Kim himself [1993, 1998]), regularly talk of events as instances of properties—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.

(7) An event's constitutive property can no more be viewed as a property *of* it than its constitutive object can be viewed as a property of it. The claim that *P* is a constitutive

property of *e* entails, not that *P* is a property of *e*, but rather, that being an exemplifying of *P* is a property of *e*. Thus, for example, the claim that the property, *firing*, is a constitutive property of the event which is Joe's firing a gun at *t* entails, not that *firing* is a property of that event, but rather, that being an exemplifying of the property, *firing*, is a property of that event.

(8) The exposition of the PEA here is based on this work of Kim's (see especially Kim 1973, 1976). According to Kim, although the first condition 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.

(9) More precisely, objects that are *minimal subjects* of events, since, as stated in the text, we prefer the version of the account developed by Lawrence Lombard (1986). According to this, an object, *x*, is the minimal subject of an event *e* if it is the minimally involved subject of *e*, where the notions of an object's involvement and minimal involvement in an event are defined as follows:

> If *x* is any object, *e* is any event, and *t* is a time, then *x* is involved in *e* at *t* if and only if it is the case that if *e* occurs (or is occurring) at *t*, then *x* changes (or is changing) at *t*, and a change in *x* at *t* is identical with *e* at *t*; and
>
> If *x* is any object, *e* is any event, and *t* is a time, then *x* is the minimally involved subject of *e* at *t* if and only if (a) *x* is involved in *e* at *t*, and (b) *x* is the smallest object which is such that a change in *x* at *t* is identical with *e* at *t*. (Lombard 1986: 122–3)

For more on the details of Lombard's version of the PEA, and his reasons for distinguishing subjects from minimal subjects of events, see Lombard (1986). That mental and physical events might have different subjects—mental ones having, say, persons, and physical ones having, say, brains—does not preclude identity between mental and physical events on the PEA, since the distinction between subjects and minimal subjects allows for the possibility that persons are not minimal subjects of mental events. For more on this, see Macdonald (1989).

(10) See Lombard (1986), who points out that the view that events may have more than one constitutive property is *not* inconsistent with the existence and identity conditions of events as stated by the PEA and formulated in the text here:

> 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. (Lombard 1986: 54–5)

(11) Kim claims on behalf of the PEA that both mental properties (of persons) and physical properties (of persons) are constitutive properties of events, and, in his early work (Kim 1972) he concludes that token identity theories of the mind#body relation are false, on the grounds that mental properties are not identical with physical ones, but the PEA is *not* committed to this conclusion, for two reasons. First, it requires that an event cannot have more than one constitutive property, and the PEA need not be committed to this (cf. our discussion of Lombard's version of the PEA, which rejects it). Second, even if one does suppose it, one might claim—as we do—that mental properties of persons supervene on the constitutive properties of physical events, and so are not constitutive properties of those events (see subsequent pages of the text). We prefer this way of reconciling the PEA with token event identity because we think that a proper physicalism must be committed not just to an ontology of physical events, but also to providing an explanation of the relation between mental and physical *properties* which shows them to be, if not physical, not worryingly non#physical. The first way of reconciling the PEA with token event identity leaves the question of the relation between mental and physical properties completely open.

(12) We have re#labelled Kim's 'P*' as 'P2' for consistency with what has gone before.

(13) 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, on the one hand, and talk of properties and causal relevance, on the other.

(14) Where Kim and others would use 'higher#order/lower#order' terminology we use the terminology of 'higher#level/lower#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 pain, or dispositional properties like being soluble. But this is quite different from the contemporary logician's usage (though similar to Russell's and Ramsey's). In contemporary terms, 'being soluble', 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, '$\#F(Fx \ \& \ \#y(Fy \ \& \ y$ is placed in a relevant liquid# y dissolves)', where we have a second#order quantifier, '$\#F$', which ranges over all properties, including being soluble (just as the bound variable in '$\iota x(\#y)(x$l$y \ \#$ x is taller than $y)$' impredicatively specifies the tallest person.

(15) Pettit places three conditions on a property's being a 'programming property':

1. (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. (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. (3.) The lower#order properties associated with the actual instantiation of the higher#order property do in fact produce E. (Pettit 1993: 37)

(16) Thus, he says, 'The key question is whether the emergent property instance H is realized and focusing upon the non#causal nature of the determination exerted by the instance of H shows that it is indeed realized in this case.The central point that we need to emphasize is that H is *not causing* P1 to contribute certain powers' (Gillett 2006a: 282).

(17) So, for example, despite the emphasis placed in the quote in note 16 on the distinctness of the emergent property instance and the realizer property instance, and despite his claim that 'It should also be marked that under Alexander's concept it is plausibly property instances that are emergent' (2006a: 292), he says, 'an emergent property is identical to a combination of such properties which is itself realized.But

the emergent property is not metaphysically "nothing but" the realizers. For no *merely* microphysical set of properties by themselves account for the causal powers contributed by this combination of properties to individuals. Only by taking the lower level properties to be "new" in realizing H, i.e., the emergent property instance, can we account for certain of the powers of the new "constellation" of microphysical properties. Thus, in such a case, a realized property like H, albeit one identical to some combination of lower level properties and relations, can be a necessary member of a set of properties that are only jointly sufficient for contributing a certain causal power, in Cx, to an individual such as *a*1′ (Gillett 2006a: 282).

(18) According to him, the program account 'would make sense of the causal relevance of intentional states, improving considerably on the most influential current alternative: the story that gives them relevance through construing them as identical with electronic or neural states. That story had the defect that it would make a state like the belief that p causally relevant but 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. . . . Consider the case where the belief that p happens to have the same realiser state as the belief that q. Both the belief that p and the belief that q will have the same neural or electronic character but the program account explains how the belief that p may be causally relevant to an action of A#ing in a way in which the belief that q is quite irrelevant (Macdonald and Macdonald 1986)′ (Pettit 1993: 38). We take it that the first accusation, that 'the belief that p would be causally relevant in virtue of being a neural state′ is the accusation that the co#instantiation model, by identifying instances of these two properties, makes the property, being a belief that p, causally *efficacious* by being co#instantiated with the property of being a certain neural state. So it does, but since our solution to the problem of mental causation distinguishes causal efficacy from causal relevance, the former being a necessary but not sufficient condition for the latter, the co#instantiation model does not thereby establish that mental properties are causally relevant only 'in virtue′ of the causal relevance of their realizing properties. Further, given that both we and Pettit accept that mental (and higher#level) properties supervene on their realizers, the possibility that mental properties could be realized by the same physical ones is ruled out. So the second part of the quote again must refer to instances, not properties, and causal efficacy, not causal relevance. But given the co#instantiation model, causal efficacy of the mental property is secured and cannot be eclipsed by that of its physical realizer.

(19) Again, where Pettit and others would speak of mental properties as higher#order properties, we would use the term 'higher#level′. See note 14.

(20) In this Pettit's example differs from the kind of case that Gillett's view is concerned with, a case where the emergent property instance is mereologically related in a part#

whole way, to all of the instances of the lower#level, microphysical properties that realize it. See Gillett 2006b.

(*) This section draws on work done in Macdonald and Macdonald 2007.