

EMERGENCE, DOWNWARDS CAUSATION AND THE COMPLETENESS OF PHYSICS

BY DAVID YATES

The ‘completeness of physics’ is the key premise in the causal argument for physicalism. Standard formulations of it fail to rule out emergent downwards causation. I argue that it must do this if it is to feature in a valid causal argument for physicalism. Drawing on the notion of conferring causal power, I formulate a suitable principle, ‘strong completeness’. I investigate the metaphysical implications of distinguishing this principle from emergent downwards causation, and I argue that categoricist accounts of properties are better equipped to sustain the distinction than dispositional essentialist accounts. Finally, I argue that the additional evidence needed for strong completeness renders the causal argument otiose for any properties amenable to scientific reduction.

The ‘completeness of physics’ is (roughly) the claim that every physical effect has a sufficient physical cause. Various precisifications have been suggested, but the common idea is that all the causal work needed for physical events to occur (however ‘causal work’ ought to be understood) is done by physical causes. Completeness is the key premise in the causal (or overdetermination) argument for physicalism; this is a general argument form which can be used to argue for physicalism about any domain Φ of causes with physical effects.¹

- (a) Φ -causes sometimes have physical effects
- (b) Every physical effect has a sufficient physical cause
- (c) Effects do not usually have more than one distinct sufficient cause.

¹ Versions occur in D.K. Lewis, ‘An Argument for the Identity Theory’, *Journal of Philosophy*, 63 (1966), pp. 17–25, and ‘Psychophysical and Theoretical Identifications’, *Australasian Journal of Philosophy*, 50 (1972), pp. 249–58; J. Hopkins, ‘Mental States, Natural Kinds and Psycho-Physical Laws’, *Proceedings of the Aristotelian Society*, Supp. Vol. 52 (1978), pp. 221–36; D. Papineau, ‘Why Supervenience?’, *Analysis*, 50 (1990), pp. 66–71, and ‘Mind the Gap’, in J. Tomberlin (ed.), *Philosophical Perspectives 12: Language, Mind and Ontology* (Oxford: Blackwell, 1998), pp. 373–89; J. Kim, *Mind in a Physical World* (MIT Press, 1998), and ‘Blocking Causal Drainage and Other Maintenance Chores with Mental Causation’, *Philosophy and Phenomenological Research*, 67 (2003), pp. 151–76. That all versions depend on completeness is argued in T. Crane, ‘Mental Causation’, *Proceedings of the Aristotelian Society*, Supp. Vol. 69 (1995), pp. 211–36.

Let the domain be that of mental causes, and let m be a mental cause of physical effect e . From (b) we know that e has a sufficient physical cause p . But from (c) we can conclude that m and p are not distinct. This argument takes a lot of unpacking. Must mental causes be efficacious *in the same way* as physical causes for the argument to go through?² Does the argument require that mental and physical causes must be ‘in competition’ for *the very same effects*?³ Would non-identical, but metaphysically related, causes of the same effect, count as *overdetermining* causes? Relatedly, should the conclusion of the argument, that the causes in question are not distinct, be understood in terms of strict identity, or some form of metaphysical dependence?⁴

I shall not address such issues in the present paper. My main aims in this paper are (i) to show that standard formulations of the completeness of physics are too weak to feature in a valid causal argument for physicalism; (ii) to formulate a suitably strengthened completeness thesis; and (iii) to investigate the metaphysical and epistemological implications of strengthening completeness in this way. I follow Frank Jackson in defining physicalism as the claim that any possible world that is a minimal physical duplicate of the actual world is a duplicate *simpliciter*.⁵ Space prevents a detailed treatment, but the idea is this: if physicalism is true, then if God wants to make an exact duplicate of the actual world at a given time, all he has to do is create a world indistinguishable from actuality at that time in terms of the distribution of basic physical particulars and properties, and make the laws of physics true. The rest he gets free. Physicalism about the mind I take to be the claim that any minimal physical duplicate of the actual world is also a mental duplicate. In addition, for ease of exposition, I shall take the causes mentioned in the premises of the causal argument to be property-instances. I conceive these as ‘Kim events’, which are structured particulars involving objects (or more generally, substances) possessing properties at times.⁶ I hold that properties confer causal powers upon the particulars that instantiate them, and that property-instances cause their effects in virtue of being instances of properties that confer the power to cause those effects. (I shall sometimes speak of causes as possessing powers.) When, in what follows, I talk of properties as causes, I should be taken to be referring to their instances.

² See Crane, ‘Mental Causation’, pp. 17–21, for discussion.

³ See S. Sturgeon, ‘Physicalism and Overdetermination’, *Mind*, 107 (1998), pp. 411–32.

⁴ See T. Sider, ‘What’s So Bad about Overdetermination?’, *Philosophy and Phenomenological Research*, 67 (2003), pp. 719–26.

⁵ F. Jackson, ‘Armchair Metaphysics’, in J. Hawthorne and M. Michael (eds), *Philosophy in Mind* (Dordrecht: Kluwer, 1994), pp. 23–52.

⁶ Kim, ‘Events as Property Exemplifications’, repr. in his *Supervenience and Mind: Selected Philosophical Essays* (Cambridge UP, 1993), pp. 33–52.

I. EMERGENT PROPERTIES AND DOWNWARDS CAUSATION

My purpose in this section is to formulate an emergentist position, to which I shall appeal in the remainder of the paper, in order both to define the completeness of physics and to determine what would count as evidence for it. I should point out that I do not attribute emergentism as characterized here to any other philosophers, although the position I describe owes much to (and is not dissimilar to the emergentism of) C.D. Broad. I do not endorse my emergentism: my arguments will go through provided it is (i) internally consistent, and (ii) inconsistent with the completeness of physics. As I shall conceive it, emergentism is the conjunction of a metaphysical thesis, concerning the relationship between physical and emergent properties, and a thesis concerning the causal powers of emergent properties.

First, the metaphysical thesis. Emergentism has much in common with physicalism. It is *properties* that are emergent: there is no emergent substance. Emergentism thus shares with physicalism the thesis of physical monism. Further, emergentism shares with physicalism the view that the emergent properties are determined by, and supervenient upon, physical properties. Emergent properties are properties of certain physical structures, synchronically determined according to what Broad terms ‘trans-ordinal laws’ whose antecedents denote complex physical properties of such structures, and whose consequents denote emergent properties.⁷ These laws are ‘unique and ultimate’ laws true in some physically possible worlds, not in others. To put these ideas together, emergent properties are instantiated by physical particulars with a sufficient level of complexity, and are determined by complex physical properties of those particulars according to laws which hold (or do not hold) independently of the laws of physics. Emergent properties are nomologically necessitated by physical properties; but, given emergentism, some laws of nature are not true in all physically possible worlds. The metaphysical component of emergentism entails that emergent properties, if there are any, would falsify physicalism as I am defining it. Minimal physical duplicates of a world do not duplicate the trans-ordinal laws, and so no emergent properties are instantiated at such worlds. The present characterization leaves it open whether the determination of emergent properties by physical properties is causal or non-causal. For ease of exposition in later sections, I shall suppose it to be causal (supposing also that synchronic causation is metaphysically unproblematic). Little of import

⁷ See C.D. Broad, *The Mind and its Place in Nature* (London: Routledge & Kegan Paul, 1925), pp. 77–80, for discussion of trans-ordinal laws.

turns on this assumption, and my central arguments will go through *mutatis mutandis* if the determination is non-causal.

Secondly, the causal thesis. Emergent properties exert a downwards influence on the physical world. Let S be a physical emergence base for an emergent property F, which emerges in virtue of a physical property P of S. S is the constitutive object for two events, a P-instance and an F-instance, with the latter emerging from the former. The causal thesis is just the claim that F confers causal powers upon S not conferred by P. Certain properties of the components of S will be, to borrow C. Lloyd Morgan's word, 'additive'.⁸ Their inertial masses, for instance, can be summed to find the inertial mass of S, according to the principle of scalar addition for inertial mass. We can calculate how much force it takes to accelerate S at some rate, by adding the forces it takes to accelerate its components at that rate. But if S is the base for an emergent property, then not *all* of its causal powers are like that. If inertial mass were an emergent property of S (which of course it is not), then we would get the wrong result if we tried to find the mass of S by summing the masses of its components. I am not claiming that emergent properties somehow *take over* from the physical properties they emerge from. The physical properties of S still contribute causal powers to S; the (physically) unexpected effects are due to *extra* powers conferred by the emergent properties. According to this view, emergent properties *combine* with their physical base properties to cause certain physical effects, and these effects would not occur but for the causal powers of the emergents.

I shall now argue that emergentism as characterized above is inconsistent with the completeness of physics.

Plausibly, any duplicate *simpliciter* of a deterministic world w at some time t is a duplicate *simpliciter* of w at any $t^+ > t$. Let w_e be a deterministic world at which an emergent property F is sometimes instantiated, and suppose that x is F at t_e . Further, suppose that the causal powers contributed by F to x at t_e are powers without which some physical event at $t_e^+ > t_e$ would not have occurred. By the metaphysical thesis of emergentism, we know that F will not be instantiated at a minimal physical duplicate w of w_e at t_e . But, combining the metaphysical and causal theses, we get the conclusion that a minimal physical duplicate of w_e at t_e will not minimally duplicate w_e at t_e^+ . This seems to be a consequence that no proponent of completeness would accept: surely, if w and w_e diverge between t_e and t_e^+ in virtue of the instantiation of a non-physical property F at w_e , then at w_e some of the causal work required for certain physical effects to happen is being done by F? It intuitively appears, then, that the completeness of physics is false at w_e .

⁸ See C.L. Morgan, *Emergent Evolution* (London: Williams & Norgate, 1923), pp. 2–3.

But this is too quick. The term ‘completeness of physics’ is not one for which one can test definitions against intuitions.⁹ Rather, it is a technical ‘term of art’ introduced to serve a particular role in philosophical argument. A significant part of the role in question, however, is just its role in the causal argument – and this paves the way for a simple argument to the effect that if completeness is true, then emergentism is false. However completeness is to be defined, it must be so defined that it does the philosophical work it was introduced to do in the first place. Suppose for the sake of argument that there is a valid causal argument for physicalism. I shall take the completeness of physics to be whatever principle must be combined with the other two premises (*viz* that the properties in the target domain have physical effects, and the ban on widespread overdetermination) in order to yield this argument. Now suppose for *reductio* that this completeness principle is true at w_e . Again let F be an emergent property whose instantiation at w_e makes a difference to the physical goings-on there. Suppose further that there is no widespread overdetermination at w_e . Since completeness (by hypothesis) is true at w_e , we have a valid causal argument for physicalism about F, which is by definition a non-physical property. The causal argument must be unsound at w_e , so, contrary to the supposition, completeness is *not* true at w_e .

In the above argument I assume that there is no overdetermination at w_e . This should not be taken to mean that I take it to be necessarily true, or even true, that there is no overdetermination. My argument will go through provided it is not necessarily false that there is no overdetermination. The key move in the argument is the claim that completeness is the principle which, in conjunction with the other two premises of the causal argument, yields a valid argument for physicalism. Assuming completeness to be consistent with emergentism, then at those worlds where there is no overdetermination, such as w_e , there is a valid argument for physicalism about emergent properties such as F. It follows that completeness cannot be consistent with emergentism. Of course, one might simply insist that F is really a peculiar kind of *physical* property. As others have pointed out, however, this sort of strategy renders completeness true by definition, and of no use in arguing for a non-trivial brand of physicalism.¹⁰ However completeness is to be defined, then, it must be defined in such a way as to rule out emergence. As I shall show, supplying such a definition is far from straightforward. I turn now to the task of defining completeness.

⁹ I owe this point to an anonymous referee.

¹⁰ T. Crane and D.H. Mellor, ‘There is no Question of Physicalism’, *Mind*, 99 (1990), pp. 185–206; Crane, ‘Why Indeed? Papineau on Supervenience’, *Analysis*, 51 (1991), pp. 32–7.

II. DEFINING THE COMPLETENESS OF PHYSICS

The purpose of this section is to show that standard definitions of completeness do not rule out emergence, and to suggest a strengthened thesis which does. The completeness thesis I shall suggest will be a conjunction of the standard principle with an additional thesis which in effect amounts to a denial that there is any downwards causation. I shall appeal to the distinction between strong and weak completeness theses in order to rebut Jaegwon Kim's argument against the coherence of emergentism.¹¹ I shall refer to the following thesis as 'weak completeness':

CPW. Every physical effect that has a sufficient cause has a sufficient physical cause.

(Readers worried about indeterministic causation may understand a sufficient cause to be a cause that suffices for the *chances* of its effects.) As Jonathan Lowe argues, (CPW) requires only that physical effects have sufficient physical causes *somewhere* in their aetiology.¹² Let a physical event P be a sufficient cause of some non-physical event N, which in turn sufficiently causes physical effect E. By the transitivity of causation, E has a sufficient physical cause P. An obvious reformulation introduces reference to times:

CPW t . Every physical effect that has a sufficient cause at t has a sufficient physical cause at t .

This is clearly an improvement, but does not go far enough. As Lowe points out, (CPW t) neglects the possibility of forms of emergentism which allow *synchronic causation*. Although Lowe does not explicitly subscribe to any particular form of emergentism, a moment's reflection will show that (CPW t) is consistent with emergentism as defined in §I. Let F emerge from base state S at t , in virtue of a complex physical property P. Suppose P and F together confer upon S the necessary powers to cause diachronically some effect E. E has a sufficient cause at t , *viz* the P- and F-instances. However, since the P-instance is causally sufficient for the F-instance, it follows (again by transitivity of causation) that E has a sufficient *physical* cause at t (S's being P) despite the downwards causal influence of F. Completeness ought to rule out downwards causation, but (CPW t) fails to do so. I shall term the following modification 'strong completeness':

¹¹ Kim, 'Making Sense of Emergence', *Philosophical Studies*, 95 (1999), pp. 1–34.

¹² E.J. Lowe, 'Causal Closure Principles and Emergentism', *Philosophy*, 75 (2000), pp. 571–85.

CPS. Every physical effect that has a sufficient cause at t has a complete sufficient physical cause at t .

The trouble with (CPW t) is that the physical effects of emergent properties have sufficient physical causes. Still, it seems clear that those causes are not doing all the causal work required for their effects. (CPS) adds in the proviso that in addition to sufficiency, the physical causes of physical effects are also *complete*. Intuitively, if emergent properties are doing some of the causal work required for an effect E, then their emergence physical bases will not be complete causes of E. But how should 'complete' be defined? It is natural at this point to appeal to the laws of physics. David Papineau defines the completeness of physics as follows: 'all physical events are determined ... entirely by prior physical events according to physical laws', and 'All physical effects have complete physical causes ("complete" in the sense that those causes on their own suffice by physical law to fix the chances of those effects)'.¹³ Again suppose an emergent property F has base property P, and their combined effect is E. The trans-ordinal laws relating P to F are not *physical* laws. But since F confers powers necessary for the occurrence of E, it follows that P does not cause E 'according to physical laws', and so is not a *complete* cause.

As others have argued, however, it is far from clear that the laws of physics fail to apply in cases where an emergent property is causally active. The physical emergence base properties still confer the causal powers they would in a physically possible world with no emergent properties. But then if 'according to physical law' is understood merely in terms of consistency with the laws of physics, it will be of no use in the present context. According to Terence Horgan,

Does [downwards causation] mean that the laws of physics are abrogated when emergent properties are instantiated? According to the emergentists, no. For the laws of physics *do not actually assert that physical forces are always the only operative forces in a physical system*.¹⁴

If this is true, then the operation of emergent properties is, at least in one sense, in accordance with, or according to, physical laws. Interestingly, Papineau agrees, arguing convincingly that emergent properties do not violate the laws of physics, including conservation of energy.¹⁵ It seems that

¹³ Papineau, 'Why Supervenience?', p. 67, and *Philosophical Naturalism* (Oxford: Blackwell, 1993), p. 22, respectively.

¹⁴ T. Horgan, 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World', *Mind*, 102 (1993), pp. 555–86, at p. 557 (fn. 4); my italics.

¹⁵ Papineau, 'The Rise of Physicalism', in C. Gillett and B. Loewer (eds), *Physicalism and its Discontents* (Cambridge UP, 2001), pp. 3–36. See also B. McLaughlin, 'The Rise and Fall of British Emergentism', in A. Beckermann, H. Flohr and J. Kim (eds), *Emergence or Reduction: Essays on the Prospects of Non-reductive Physicalism* (Berlin: De Gruyter, 2001), pp. 49–93.

Papineau means something stronger by ‘according to physical laws’ than mere *consistency* with physical laws; I shall return to this issue in §III.

Emergent properties do causal work not done by their physical base properties. As I conceive it, the causal work of a property consists in conferring causal powers upon particulars that instantiate it. Why not then define ‘complete’ in (CPS) explicitly in terms of causal power conferring? I shall represent a property-instance as an ordered triple $[x, P, t]$, and define a complete sufficient cause as follows:

Comp. A property-instance $[x, P, t]$ is a complete sufficient cause of an effect E if and only if (i) $[x, P, t]$ is a sufficient cause of E; (ii) all the powers needed to cause E are conferred upon x by P.

Given (Comp), (CPS) claims that every physical effect E has a sufficient physical cause whose physical properties confer all the powers needed to cause E, and seems to be inconsistent with emergentism, which claims that some of the powers to cause certain physical effects are *not* conferred by physical properties of their causes. In the remainder of this section, I shall show how the distinction between (CPS) and (CPW t) enables a rebuttal of Jaegwon Kim’s argument that emergentism is incoherent.

Emergentism has a metaphysical component and a causal component. According to the first, emergent properties supervene on their emergence base properties; according to the second, they confer causal powers over and above those conferred by their bases. Kim argues that these two components are inconsistent:

... if an emergent, F, emerges from basal condition P, why can’t P displace F as a cause of any putative effect of F?... Now we are faced with P’s threat to pre-empt F’s status as a cause of E.... For if causation is understood as nomological (law-based) sufficiency, P, as F’s emergence base, is nomologically sufficient for it, and F, as E’s cause, is nomologically sufficient for E. Hence, P is nomologically sufficient for E and hence qualifies as its cause.... This appears to make the emergent property F otiose and dispensable as a cause of E.¹⁶

Kim’s argument that P displaces F as cause of E is based on the fact that P is nomologically sufficient for E. One advantage of defining emergence in terms of synchronic causation is that I can grant Kim that P is causally sufficient for E without commitment to a nomic sufficiency analysis of causation. That P is a sufficient cause of E follows from the nature of emergence and the transitivity of causation. P is a sufficient cause of F, (P&F) is a sufficient cause of E, so P is a sufficient cause of E. Clearly, Kim would not allow that earlier causes in a chain of sufficient causation to E displace more

¹⁶ Kim, ‘Making Sense of Emergence’, p. 32 (adjusted for typographic consistency).

proximal causes; what is driving Kim's argument, then, is the fact that S and F are equally proximal causes of E. This in turn suggests to me that Kim (not uncommonly) takes (CPW t) to define the completeness of physics. He has something like the following argument in mind:

1. The causal thesis of emergentism entails the incompleteness of physics
2. The conjunction of the causal and metaphysical theses entails the completeness of physics
3. The completeness of physics entails the falsity of the causal thesis
4. Therefore if the metaphysical thesis is true, then if the causal thesis is true, the causal thesis is false.

Thus if emergent properties supervene on the physical, then they do not exert a downwards causal influence on the physical. The argument looks promising, but equivocates between (CPS) and (CPW t). What underlies (1), and of course its contrapositive (3), is, I submit, an intuitive grasp of what it means for physics to be causally complete. For instance, it seems clear that the following conditional is true: if physics is causally complete, then any causal powers a cause needs in order to produce a physical effect are conferred by its physical properties alone. Conversely, if the occurrence of an effect requires the cause to have powers in addition to those conferred by its physical properties, then physics is not complete. As I argued in §I, these intuitions can be supported by considering the philosophical work which physicalists expect completeness to do. These considerations entail that (CPW t) is too weak to define completeness, and that a stronger thesis is needed, such as (CPS). Thus in (1), completeness must be read strongly, as in (CPS); if it is read in this way, however, (2) is false. Emergentism does not entail (CPS). Emergentism *does* entail (CPW t), but read in this way, (1) and (3) are false, for the causal thesis is consistent with (CPW t).

The distinction between (CPS) and (CPW t) depends on the possibility that a sufficient cause of some effect E may not be a *complete* cause of E. However, there is a problem. On at least one account of the relationship between properties, powers and effects, the base property P of an emergent property F will *confer* the supposedly novel powers of F. But if this is so, then it seems Kim is right to doubt the coherence of emergentism, for emergent properties by definition confer powers *not* conferred by their base properties. What is needed is an account of these relations which divides up causal powers, in a principled way, between emergent properties and their bases.

III. PROPERTIES, POWERS AND EFFECTS

How should we understand the relations between properties and the powers they confer, and between those powers and the effects they are powers to cause? Jessica Wilson suggests this criterion for ‘causal power bestowal’:

A property P bestows a causal power C(K,E) [if and only if] instances of P, in circumstances K, are nomologically sufficient for instances of E, and the holding of K alone is not nomologically sufficient for instances of E.¹⁷

As Wilson argues, according to the above criterion, emergence base properties will confer the putatively novel powers of the emergents. Let property-instance $[x,P,t]$ be a base for an emergent property-instance $[x,F,t]$, and suppose for *reductio* that F confers upon x the power, not conferred by P, to cause (an instance of) E. On the current understanding of conferring, this means that $[x,F,t]$ must be nomologically sufficient for E (I set aside for the moment the role of K). $[x,P,t]$, as F’s emergence base, is nomologically sufficient for $[x,F,t]$. By transitivity of sufficiency, it follows that $[x,P,t]$ is nomologically sufficient for E. But on the present conception of conferring, this entails that P confers upon x any powers conferred by F, in which case F does not confer any *novel* powers at all. On the present account of conferring, then, emergentism is incoherent.

At this point, there are the following options: (i) redefine causal power conferring, to preserve the coherence of emergentism; (ii) appeal to something other than conferring to distinguish the powers of emergent properties from those of their bases; or (iii) argue from the present conception of conferring to the incoherence of emergence. Wilson opts for (ii), after arguing that (i) will not work, whereas (iii) echoes Kim’s argument against emergentism. In the remainder of this section, I shall defend (i). To this end I shall divide the relationship between properties, powers and effects into two: (a) a relationship between properties and the powers they confer; (b) a relationship between those powers and the effects that they are powers to cause. I shall understand the conferring relationship as follows, leaving it open at this point what force the modal operator has:

Conf. A property P confers a power ϕ on a particular x if and only if, necessarily, if x is P then x has ϕ .

¹⁷ J. Wilson, ‘Causal Powers, Forces and Superdupervenience’, *Grazer Philosophische Studien*, 63 (2002), pp. 53–77, at p. 62.

The importance of this step consists in separating the power-conferring relation from the relation between powers and effects. How should the latter be understood? Jerry Fodor proposes a condition for identity of causal powers that will help:

Roughly, our biceps have the *same* causal powers if the following is true: *for any thing x and any [nomologically possible] context C , if you can lift x in C , then so can I; and if I can lift x in C , then so can you.*¹⁸

I follow Denis Walsh in thinking of a context as a set of properties, where x counts as being *in* a context if it is ‘causally connected, in an appropriate way, to an environment characterized by that context’.¹⁹ As things stand, Fodor’s condition has an air of circularity. Fodor does not say how we are to understand ‘can’, but a natural way to flesh it out is in terms of causal powers. Then Fodor’s claim amounts to this: that two biceps A and B have the same causal powers if, for any x , A has the power to lift x if and only if B does. This is hardly informative. But Fodor (p. 44) goes on to say ‘Identity of causal powers is identity of causal *consequences* across nomologically possible contexts’ (my italics). This is rather more informative, and suggests an identity criterion for causal powers based on identity of effects across nomologically possible contexts (which I label ‘Fodor’s criterion’, without thereby attributing it to Fodor):

FC. For any x and y , x has the same causal powers as y if and only if for any nomologically possible context C and any effect E, x causes E in C if and only if y causes E in C.

(FC) entails that the relationship between powers and contexts, on the one hand, and effects, on the other, is nomologically necessary. The causal consequences of an x cannot vary, holding both the powers of x and the context fixed – otherwise it would be nomologically possible for some x and y to have the same causal powers, be in the same context, and yet have different effects, which violates (FC). I shall refer to this as the ‘power-effect principle’, and define it as follows:

PEP. For any x , if x has the power to cause some effect E in a context C, then it is nomologically necessary that if x is in C, then E occurs.

It ought to be clear how (Conf) and (PEP) combine to generate the problem outlined above, for together they entail Wilson’s criterion, provided the necessity in (Conf) is understood as nomological. According to

¹⁸ J. Fodor, *Psychosemantics* (MIT Press, 1989), p. 35.

¹⁹ D.M. Walsh, ‘Alternative Individualism’, *Philosophy of Science*, 66 (1999), pp. 628–48, at p. 633.

emergentism, instantiation of their physical base properties alone is sufficient to cause the emergents. In essence, the ‘circumstances’ are built into the definition of a base property: emergence bases have the power to cause emergents without the need for a context over and above the instantiation of the base properties. I shall call this feature of emergence ‘context insensitivity’. Let $[x, P, t]$ be an emergence base, and allow for the moment that the strength of the necessity in (Conf) is nomological. Base property P confers upon x the power to cause x to have emergent property F. Thus by (PEP) and the context insensitivity of emergentism, it is nomologically necessary that if x is P, then x is F. Suppose for *reductio* that F confers a novel power ϕ on x . Again by (Conf) it will be nomologically necessary that if x is F, then x has ϕ . But it follows by transitivity of sufficiency that it is nomologically necessary that if x is P, then x has ϕ , which by (Conf) entails that P confers ϕ upon x .

In the remainder of this section I shall consider ways of revising (Conf) and (PEP) in order to distinguish powers conferred by emergent properties from those conferred by their bases. The methods I shall consider are (1) modifying (Conf) by changing the strength of the modality involved; (2) modifying (PEP) to weaken the connection between powers and the effects they are powers to cause. I shall consider these alternatives in the light of what I take to be the two main realist approaches to causal powers: the *categoricalist* (or ‘DTA’) account, endorsed by Fred Dretske, Michael Tooley and David Armstrong;²⁰ and the *dispositional essentialist* account (hereafter ‘dispositionalism’), endorsed by philosophers such as Sydney Shoemaker and Alexander Bird.²¹ Only categoricalists, I shall argue, can endorse (1); dispositionalists are limited to (2). This result leaves dispositionalists with some awkward problems which categoricalists do not have.

According to categoricalism, universals lack any essential modal character. The causal powers conferred by universals are determined by contingent necessitation relations that hold between them. For first-order universals F and G, a law of nature is the obtaining of a second-order necessitation relation N between them, N(F,G). Particular causal relations, on this view, are grounded in N-related universals. According to dispositionalism, it is in the essential natures of properties to contribute the causal powers they do; the laws of nature are a consequence of the natures of properties. I need not go into details. For my present purposes, the key difference between categoricalism and dispositionalism is the strength of the

²⁰ See, for instance, D.M. Armstrong, *What is a Law of Nature?* (Cambridge UP, 1983).

²¹ S. Shoemaker, ‘Causality and Properties’, in P. van Inwagen (ed.), *Time and Cause* (Dordrecht: Reidel, 1980), pp. 109–35; A. Bird, ‘Strong Necessitarianism: the Nomological Identity of Possible Worlds’, *Ratio*, 17 (2004), pp. 256–76.

necessity in (Conf). Dispositionalism holds that the powers conferred by a property are essential to it, and so the necessity in (Conf) will be metaphysical. Categoricalism, on the other hand, holds that the powers conferred by a property depend on metaphysically contingent laws of nature, so it is nomologically necessary that properties confer the powers they confer. My problem may now be restated as follows. Given categoricalism, the necessity in (Conf) is nomological. But as I said above, together with (PEP), this entails that emergence base properties confer the putatively novel powers of emergents. Given dispositionalism, (Conf) holds with metaphysical necessity. But dispositionalism entails that the laws of nature are metaphysically necessary, so that there is no difference between nomological and metaphysical necessity, and the necessity in (PEP) is also metaphysical. It therefore follows, *mutatis mutandis*, that any putatively novel powers attributed to emergent properties are conferred by their bases.

First, a categoricalist response. Since the powers properties confer are determined by the laws of nature, and these laws define the set of nomologically possible worlds, a property will confer the same powers in every nomologically possible world. Assuming emergentism is true, not all the fundamental laws of nature are laws of physics: there are additional laws governing emergence and the causal powers that emergent properties confer. For categoricalists, properties are essentially non-modal; so for a property to be physical is *for its behaviour to be governed by the laws of physics*. All the natural laws that govern a property will determine what powers it confers *simpliciter*. But now one can distinguish the powers an emergence base property confers *simpliciter* from those it confers *qua* physical. Given emergentism, the latter will be a subset of the former – but which subset? Wilson (pp. 69–74) argues that we cannot define this set in terms of a modified account of conferring. Rather, she appeals to a primitive ‘grounding’ of causal powers in fundamental forces, and argues that the powers conferred by a property *qua* physical are those of its powers *simpliciter* that are grounded in fundamental physical forces. This approach involves reification of forces, and a primitive relation, *grounding*, which seems at least as obscure as the ‘conferring *qua*’ relation it is supposed to explain. There is an alternative, however, which Wilson does not consider. The distinction between physical and non-physical *laws* can be used to ground the distinction between conferring *qua* physical and conferring *simpliciter*.

Let Φ be a set of laws governing the behaviour of a categorical property, and understand a ‘ Φ -property’ to be a property so governed; correspondingly, Φ -necessity will be truth in all possible worlds where the laws in Φ are true. Here is a modification of (Conf):

Conf Φ . A Φ -property P confers a power ϕ *qua* Φ -property if and only if it is Φ -necessary that if x is P then x has ϕ .

If Φ = the laws of nature, whatever they might be, then (Conf Φ) is equivalent to (Conf) (with nomological necessity), which will provide an account of conferring *simpliciter*. If, however, Φ = the laws of physics, then (Conf Φ) claims that a physical property P confers a power ϕ *qua* physical only if, at all physically possible worlds, if x is P then x has ϕ . Assuming the powers conferred by physical properties *qua* physical are given by (Conf Φ), it is not true that physical base properties confer the powers of emergents *qua* physical. Again suppose P is a base property, F an emergent, and ϕ the novel powers conferred by F. As before, (Conf Φ) and (PEP) entail that it is nomologically necessary that if x is P, then x has ϕ . However, since (assuming emergentism) there are fewer nomologically possible worlds than physically possible worlds, it does not follow that it is *physically* necessary that if x is P, then x has ϕ . A categorical property may confer a set ψ of powers relative to all the laws of nature, while only conferring a subset of ψ relative to the laws of physics, or of some other domain. The present account admits that P confers ϕ , but denies that P confers ϕ *qua* physical. Given (Comp), however, P will still qualify as a complete cause of any effect E of F, for (Comp) does not distinguish conferring *simpliciter* from conferring *qua* physical. According to (Comp), P need only confer all the powers required for the occurrence of E in order for $[x, P, t]$ to qualify as a complete cause of E. Fortunately, there is an obvious modification of (Comp) based on (Conf Φ):

Comp Φ . An instance $[x, P, t]$ of a Φ -property P is a complete sufficient cause of an effect E iff (i) $[x, P, t]$ is a sufficient cause of E; (ii) all the powers needed to cause E are conferred upon x by P *qua* Φ -property.

Putting together (CPS) and (Comp Φ) yields a principle which, I submit, makes considerable sense of Papineau's appeal to physical laws in defining completeness. (CPW t) holds that any physical effect E that has a sufficient cause at t has a sufficient physical cause $[x, P, t]$ at t ; (CPS) adds that $[x, P, t]$ is a complete cause of E, in the sense that all the causal powers x needs to produce E are conferred upon x by P *qua* physical. Conferring a power *qua* physical is conferring it in all possible worlds where the laws of physics are true. (Comp Φ) gives a natural way of understanding what it is for a Φ -cause to bring about an effect 'entirely according to Φ -laws', thus rationalizing Papineau's appeal to laws to explain what makes a cause complete.

It should be clear that the preceding strategy is not available to dispositionalists. The categoricalist account detailed above depends on varying the strength of the modality in (Conf) according to which set of laws we are

considering the property in question to be governed by. *Qua* physical, a property will derive its powers from the laws of physics; *qua* natural (one might say), from the laws of nature. But according to dispositionalism, it makes no sense to speak of a property conferring a power *qua* anything. Properties are individuated by the powers they confer, and not *governed* by laws at all. No modification of (Conf) is available to dispositionalists: metaphysical necessity is the only strength of modality available. They need to modify (PEP) instead. I shall consider two strategies, using (i) a probabilistic relation between powers and effects, and (ii) finks and antidotes, in each case to block the necessary connection between powers and the occurrence of the effects they are powers to cause. I shall argue that neither works.

Combining dispositionalism with an indeterministic version of (PEP) is initially attractive. Rather than treating powers as metaphysically necessitating the *occurrence* of the effects they are powers to cause, we treat them as metaphysically necessitating the *chances* of occurrence of those effects. The conferring problem for dispositionalists depends on its being metaphysically necessary that if x has base property P, then x has emergent property F. But a probabilistic (PEP) blocks the inference to this premise – what follows is only the claim that it is metaphysically necessary that if x is P, then the *probability* that x is F is $1/n$. Again supposing ϕ to be a power conferred by F, it follows that it is metaphysically necessary that if x is P, then the probability that x has ϕ is $1/n$. But from this we cannot infer from (Conf) that P confers ϕ . This solution, however, makes the coherence of emergentism depend on indeterministic causation. Why should that be? If emergentism is coherent, then there ought to be deterministic worlds with emergent properties. The present approach cannot allow for such worlds, and so is not adequate as a general account of the distinction between the powers of emergent properties and those of their bases.

The second option for dispositionalists is to hold that what Toby Handfield calls ‘space invaders’ are nomologically possible.²² These are spontaneously instantiated natural properties, possible, according to Handfield, because the laws of nature make no claims about how property-instances are brought into being. (Conservation laws seem to me to be a counter-example.) Space invaders can block the connection in (PEP) between ‘ x has the power to cause E in C and x is in C’ and ‘E occurs’, by acting either as antidotes to the power, interfering with the process that leads to the occurrence of E, or as finks, changing the intrinsic nature of x so that it no longer has the power to cause E.²³ Handfield argues that dispositionalists

²² T. Handfield, ‘Dispositional Essentialism and the Possibility of Law-Abiding Miracle’, *The Philosophical Quarterly*, 51 (2001), pp. 484–94.

²³ Lewis, ‘Finkish Dispositions’, *The Philosophical Quarterly*, 47 (1997), pp. 143–58; A. Bird, ‘Dispositions and Antidotes’, *The Philosophical Quarterly*, 48 (1998), pp. 227–34.

need space invaders in order to give a possible-worlds semantics for counterfactuals which avoids backtracking at deterministic worlds. I need not concern myself with this here. The possibility of space invaders again breaks the metaphysically necessary connection between a power and the occurrence of its effect. In essence, space-invading finks or antidotes modify (PEP) in the same way as indeterminism modifies it, but – allegedly – without *requiring* indeterminism.

A standard definition has it that determinism is true at a world w if and only if any world that has the same laws of nature as w , and is qualitatively indistinguishable from w at some time t , is also indistinguishable from w at all subsequent times. If space invaders are possible, then determinism so defined is false. Worse than that, if space invaders are consistent with *any* laws of nature, then determinism as defined above is *impossible*. Handfield anticipates such objections, and replies that the possibility of space invaders does indeed render determinism in the usual sense impossible, claiming that there is a weaker sense in which it may still be true. If it turns out that no space invaders are ever instantiated at a world w , then assuming the laws of nature at w are not probabilistic, w is ‘*de facto* deterministic’. But what could the truth of *de facto* determinism consist in? Handfield claims (p. 490) that such a world ‘would satisfy the most hard-core Laplacean variety of determinism’. Presumably the thought is that all the demon’s law-based predictions about w will come out true provided w is *de facto* a world at which no space invaders ever occur. But if this is Laplacean predictability, then why are *indeterministic* worlds not predictable? Suppose the demon simply guesses the future, relying on the probabilities; by current standards, an indeterministic world will count as predictable, provided the demon is lucky enough to guess everything right.

I conclude that categoricism has distinct advantages over dispositionalism in the present context. It is difficult to see how a dispositionalist can distinguish in a principled way between the powers conferred by an emergent property and those conferred by its emergence base. I take it in what follows that (CPS) is inconsistent with emergence, and I turn now to the question what sort of evidence would be needed to support it.

IV. EPISTEMOLOGY OF STRONG COMPLETENESS

The evidence typically cited in support of completeness theses is the progress scientists have made in finding explanations for various phenomena without using anything outside the physical domain. That scientists have made such progress is undeniable; whether their progress licenses (CPS) is another

matter. Some preliminaries. First, to reiterate: if any properties are emergent, then (CPS) is false; and if (CPS) is true, then there are no emergent properties. Thus evidence for (CPS) must be evidence against emergent downwards causation, and of course *vice versa*. Secondly, (CPS) is a universal claim, and so is violable by emergent properties *in any domain*. This creates a problem, for the domains in question are so disparate that evidence that the properties of one domain are not emergent will not count as evidence that the properties of any other domain are not emergent. (CPS) is false if there are emergent geological properties; it is also false if there are emergent mental properties. I take it we have good reason to believe that rocks are not emergent; sadly, this tells us nothing about whether psychology is. Fortunately, one can formulate domain-specific versions of (CPS), the evidence for which will be evidence that the properties in the given domain are not emergent. I shall focus on mental properties:

CPSM. Any physical effect that has a sufficient mental cause at *t* has a complete sufficient physical cause at *t*.

This principle follows *a fortiori* from (CPS), and is inconsistent with emergentism about mental properties, but is consistent with emergentism about earthquakes. A causal argument based on (CPSM) will only license physicalism *about the mind*. Domain-specific completeness theses license domain-specific physicalisms. Papineau offers an inductive argument in support of (CPSM).²⁴ Rather than couching his argument in terms of emergence, he speaks of ‘*sui generis* mental forces’, but it is clear that he would regard emergent properties as generating such forces. The argument is based on the following premises:

1. If there were emergent mental properties, then we would expect physiology to uncover them at some point
2. Physiology has made considerable progress without uncovering emergent mental properties.

The conclusion is that we have good reason to think that there are no emergent mental properties. The longer physiology goes on without uncovering such properties, the more confident we can be that it will *never* uncover any. If emergent mental properties contributed to causing physical behaviour, then I agree that there would *at some point* be a breakdown in purely physiological explanations of that behaviour. I am also happy to grant that physiology has yet to uncover emergent mental properties. However, whether or not the conclusion follows depends on whether or not

²⁴ See Papineau, ‘The Rise of Physicalism’.

the circumstances under which we would expect physiology to uncover emergent mental properties match the kind of progress it has made without uncovering such properties. In particular, the progress cited in (2) needs to be the kind of progress that we would expect physiologists not to be able to make, because of downwards causation, if there were emergent mental properties. Problematically, even if there were emergent mental properties, then there is quite a lot of progress we *would* expect physiologists to be able to make without uncovering them.

Suppose F is an emergent mental property that emerges from a neural base S in virtue of a complex neural property P of S, and S causes some physiological effect E in virtue of the combined powers conferred by P and F. As I have shown, the P-instance qualifies as a sufficient cause of E by transitivity of causation. This simple fact places severe constraints on the kind of progress that we would, and would not, expect physiologists to be able to make if there were emergent mental properties. Crucially, they *will* be able to find a sufficient neural cause P of E, and discover a law (which does not mention F) relating P-events and E-events. As Lowe puts it, the downwards causal contribution of F will be invisible to certain modes of empirical enquiry.²⁵ If physiologists are busy trying to frame laws giving nomically sufficient complex neurophysiological conditions for physiological effects like bodily movements, then they cannot be expected to uncover emergent mental properties. There is, however, a principled limit to the extent of progress in physiology if mental properties are emergent. The question whether there is emergent downwards causation from the mental to the physical comes to this: do any neural states possess causal powers over and above those conferred by their physical properties *qua* physical?

How can we decide this empirically? Suppose mental property M causes behaviour E. Take M's supervenience base S, and try to explain, in terms of laws which govern the behaviour of its component neurons, along with general composition principles of neurophysiology, how S causes E. A general composition principle is one whose applicability does not depend on how the entities it applies to are put together – for instance the principle of scalar addition for inertial masses mentioned in §I.²⁶ Either such an explanation can be given, or it cannot. If M is emergent, then applying only causal laws governing the behaviour of individual neurons, and general

²⁵ Lowe, 'Physical Causal Closure and the Invisibility of Mental Causation', in S. Walter and H. Heckmann (eds), *Physicalism and Mental Causation: the Metaphysics of Mind and Action* (Exeter: Imprint Academic, 2003), pp. 137–54.

²⁶ See A. Beckermann, 'The Perennial Problem of the Reductive Explainability of Phenomenal Consciousness – C.D. Broad on the Explanatory Gap', in T. Metzinger (ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions* (MIT Press, 2000), pp. 41–55, for useful discussion of the role of composition principles in scientific explanation.

composition principles, we would in fact expect S not to cause E, because of the downwards causal influence of M. The explanation of E fails. On the other hand, if such an explanation can be given, then we know that the causal powers of S are conferred *qua* physical by the physical properties of its neural parts, and M is not emergent. This is because the explanatory laws and principles driving the explanation will be true at all physically possible worlds. This kind of explanatory success is the only form of progress relevant to the argument from physiology.

Evidence for (CPS) is much harder to come by than evidence for (CPW t), so hard to come by, in fact, as to threaten the utility of the causal argument. The causal argument is useful only if it can establish physicalism about domains of properties we do not already know to be physical through scientific investigation. I shall argue that if we had good evidence for (CPS), we would already know that most properties were physical, without the need for additional argument. Brian McLaughlin argues that the success of quantum mechanics in providing physical explanations of chemical bonding is responsible for the decline of Broad's emergentism. Broad cited as evidence for emergence the failure of mechanistic explanations of chemical bonding; but as we now know, this failure of explanation was due to the shortcomings of pre-quantum mechanics, and not to the causal impact of emergent chemical 'configurational forces'. McLaughlin's evidence is of exactly the right kind to count against emergent chemical properties. No evidence cited by McLaughlin, however, will count as evidence against downwards causation *simpliciter*; quite rightly, McLaughlin explicitly exempts his account from application to emergent psychology.²⁷ Rather, advances in chemistry made possible by quantum mechanics suggest that there is no *chemical* emergent downwards causation:

CPSC. Any physical effect that has a sufficient chemical cause at t has a complete sufficient physical cause at t .

I agree that (CPSC) is well supported, and therefore that there is a valid causal argument for physicalism about chemicals. The question is, do we *need* any argument for physicalism about chemicals? The evidence for (CPSC) consists in lower-level explanations of chemical properties, in terms of basic physical laws, properties and composition principles. But is this not the very same evidence as also tells us that chemicals are 'nothing over and above' their physical constituents?

Suppose, with David Lewis, that scientific reductions involve some *a priori* conceptual analysis, and some empirical theorizing. The first part of the

²⁷ B. McLaughlin, 'The Rise and Fall of British Emergentism', in Beckermann, Flohr and Kim (eds), *Emergence or Reduction: Essays on the Prospects of Non-reductive Physicalism*, pp. 49–93. See p. 55, fn. 7, for the exemption.

process involves a causal–functional analysis of the concept of the property to be reduced, for instance ‘chemical bond C = the occupant of causal role R’. Then we use physics to show that the occupant of causal role R = physical property P. Then, by the transitivity of identity, it follows that C = P, and the reduction is complete.²⁸ But what does it take to show that a physical property P *occupies a causal role*? As I understand it, the role of properties in causation is to confer causal powers. On this understanding, showing that P occupies causal role R will involve showing that P confers, *qua* physical, upon a state S all the powers that S needs in order to enter into the causal relations that define R. But this is the very same evidence as is needed to justify (CPSC)! Thus if there is good evidence for (CPSC), this almost establishes a Lewisian functional reduction of chemical properties. All that remains is to fill in the *a priori* conceptual analyses which facilitate identification of the properties to be reduced with properties of the reducing theory.

Recent work by Ned Block and Robert Stalnaker casts doubt on the availability of the kind of functional reduction outlined above.²⁹ A paradigm scientific reduction is that of water to H₂O, but Block and Stalnaker point out that even assuming that our concept ‘water’ is associated *a priori* with a set of causal role specifications (which they doubt), the roles in question will not be specified in the language of the reducing theory.³⁰ Rather, the water-role will be specified in terms such as ‘colourless, tasteless, odourless liquid that boils when heated and ...’. But none of these concepts occurs in chemistry, so we need bridge principles in order to facilitate a derivation of ‘water boils at 100°C’ from our scientific theory that explains how bonds between H₂O molecules break as the average momentum of the molecules exceeds atmospheric pressure, allowing molecules to escape into the atmosphere, and so forth. These bridge principles are not *a priori*, but instead are typically supported by (e.g.) simplicity considerations, after the reducing theory is in place. I need not go into detail here. For present purposes, it will suffice to say that on Block and Stalnaker’s view of reduction, there will be more empirical work involved than on the Lewisian account. This work consists in finding *a posteriori* bridge principles to take up the slack left by the

²⁸ Lewis, ‘Psychophysical and Theoretical Identifications’, *Australasian Journal of Philosophy*, 50 (1972), pp. 249–58. That *a priori* conceptual analyses are necessary for scientific reductions is argued at length in D. Chalmers, *The Conscious Mind: in Search of a Fundamental Theory* (Oxford UP, 1996); and in D. Chalmers and F. Jackson, ‘Conceptual Analysis and Reductive Explanation’, *Philosophical Review*, 110 (2001), pp. 315–61.

²⁹ N. Block and R. Stalnaker, ‘Conceptual Analysis, Dualism and the Explanatory Gap’, *Philosophical Review*, 108 (1999), pp. 1–46.

³⁰ Similar points are made in A. Beckermann, ‘Supervenience, Emergence, and Reduction’, in Beckermann, Flohr and Kim (eds), *Emergence or Reduction: Essays on the Prospects of Non-reductive Physicalism*, pp. 94–118; and in A. Marras, ‘Emergence and Reduction: Reply to Kim’, *Synthese*, 151 (2006), pp. 561–9.

lack of *a priori* conceptual analyses. But the point now is this: once there is adequate evidence for a completeness principle such as (CPSC), supplying the bridge principles completes the reduction of chemistry to physics, again leaving nothing for the causal argument to do.

On either account of scientific reduction sketched above, the main theoretical burden in reducing a property F to the physical is the explanation of the causal role of F in terms of physical properties, laws and composition principles. This, in turn, consists in showing that the behaviour of a state S that has property F can be explained in terms of S's physical parts and their physical properties, together with general laws and composition principles which tell us how those parts will behave when put together in structures such as S. The hard part of scientifically reducing a domain of properties *just is* finding evidence for the relevant domain-specific completeness thesis. If I am right in this, then the causal argument will be otiose for any domain of properties amenable to scientific reduction. Evidence for the domain-specific completeness thesis needed for a valid causal argument about a given domain of properties gives most of what is needed to supply the kind of scientific reduction of the properties in that domain the lack of which makes it necessary to call for the causal argument in the first place.

My claim above is that the kind of evidence needed for strong completeness renders the causal argument otiose *for properties amenable to empirical reductions*. I do not, however, claim that *all* properties are amenable to such reductions. Suppose that Chalmers and Jackson are right, and (*contra* Block and Stalnaker) that *a priori* conceptual analyses are necessary for scientific property reductions. Suppose further (as is widely granted) that no such analyses are available for phenomenal concepts. It follows that the 'explanatory gap' between physical properties and phenomenal consciousness cannot be closed by empirical means. Suppose that when all the empirical evidence is in, we find that (CPS) is true. If conceptual analyses are necessary for empirical reductions, we would remain unconvinced that phenomenal consciousness is physical. However, on the assumption that instances of phenomenal properties cause physical behaviour – for example, that my instancing phenomenal painfulness causes me to wince – we can easily generate a causal argument for physicalism about phenomenal consciousness. If conscious properties confer powers upon their bearers that are also conferred (*qua* physical) by physical properties, then on pain of overdetermination, conscious properties must be physical. Physicalists thus have an argument against those who maintain that empirically irresolvable explanatory gaps entail *ontological* gaps.³¹ But they do not have a useful

³¹ See Papineau, 'Mind the Gap', for a defence of this strategy.

argument for physicalism about any domain of properties with physical effects, as is commonly supposed. The causal argument is useful only if physicalists have evidence for completeness well in advance of having evidence for physicalism; the argument aims to buy physicalism on the cheap. This evidential requirement can be met for the weak completeness theses, such as (CPW t), typically endorsed in the literature on physicalism. But such theses are of no use to physicalists, who need stronger principles such as (CPS) in order to establish physicalism. Unfortunately, (CPS) costs almost as much, in evidential terms, as physicalism itself.³²

Institute of Philosophy, University of London

³² This paper is based on research carried out while in receipt of an AHRC PhD studentship; special thanks are due to my PhD supervisor, David Papineau. Thanks also to Tim Crane, Jonathan Lowe, Eric Olson, Gabriel Segal and two anonymous referees.