FROM ***Proceedings of the Aristotelian Society Supplementary Volume*** **LXIX 1995**

The Mental Causation Debate

TIM CRANE

1. A puzzle for physicalism

This paper is about a puzzle which lies at the heart of contemporary physicalist theories of mind. On the one hand, the original motivation for physicalism was the need to explain the place of mental causation in the physical world. On the other hand, physicalists have recently come to see the explanation of mental causation as one of their major problems. But how can this be? How can it be that physicalist theories still have a problem explaining something which their physicalism was intended to explain in the first place? If physicalism is meant to be an explanation of mental causation, then why should it still face the problem of mental causation?

Disentagling this puzzle will cast light both on the recent mental causation debate and on physicalism itself. We can make a broad distinction between those forms of physicalism which identify mental and physical items and those which claim that there is some weaker relation of 'constitution' between the mental and the physical. This latter view is now the orthodox version of physicalism. I shall argue that the problem of mental causation is only a problem for this orthodox physicalism, and not for identity theories. In itself, this is not a particularly unusual claim. But I shall also argue that the real lesson of the mental causation debate is that orthodox physicalism is either unstable or unmotivated. It is unstable because (unlike the identity theories) it cannot reconcile mental causation with its other physicalist assumptions. It is unmotivated because in attempting to solve this mental causation problem, orthodox physicalism typically abandons one (or more) of the assumptions which form part of the original motivation for physicalism.

To establish this, I need to explain (a) the nature of the arguments for physicalism, (b) the problem of mental causation, and (c) the standard solutions to the problem. These three tasks will form the main substance of this paper. But first I need to make some preliminary remarks about physicalism.

2. Physicalism

By 'physicalism' I mean any theory which says that everything is physical. So if the mental exists, then the mental is physical. Since my concern here is with physicalists who accept the existence of the mental—rather than the eliminative physicalism of Quine or the Churchlands—I will use the term 'physicalism' for the more specific view that the mental is physical.

I follow physicalists in taking 'physical' to apply to anything which is the subject-matter of physics or physical science. So this includes physical particulars—such as atoms or quarks—and physical properties—such as specific masses or velocities—and entities of any other ontological category, so long as they fall under the remit of physics. (There is an important question about what 'physics' is supposed to be, but I shall have to ignore this here.) Physicalism is thus a relatively *a posteriori* thesis, whose content and justification are established by the empirical discoveries of physics.

In earlier physicalist literature, the 'is' in the phrase 'the mental is physical' was understood as the 'is' of strict identity. But recently physicalists have tended to understand the 'is' as something closer to the 'is' of constitution. To say that everything is physical in this sense is to say that everything either is a physical entity *or* is constituted by or composed of physical entities. This kind of physicalism admits that there are non-physical things—but they are exhaustively constituted by, or composed of, physical things.

The notion of *constitution* or *composition* is a notion which is most clearly applied to particular substances—for instance, to describe the relation between a substance and its parts. However, since physicalists are interested primarily in the relation between mental and physical *properties*, they will need to express their physicalism in terms of a notion of consitution/composition for properties analogous to the notion of constitution/composition for particulars.[1] Since attempts to do this are in their infancy and it is not my aim to discuss them here in detail, I will use the term 'constitution theory' just as a label for this kind of physicalism. But I don't mean to suggest by using this label that there is one fully worked-out theory which it picks out.

Physicalists have also used the idea of supervenience to express the relation between mental and physical properties. But despite the sophistication of the various attempts to define a supervenience relation adequate for the mental-physical case, it has

[1]See Yablo, 'Mental Causation'*Philosophical Review* 101, 1992, §3; Philip Pettit, 'A Definition of Physicalism' *Analysis* 53, 1992; and Jeffrey Poland *Physicalism: the Philosophical Foundations* (Oxford: Clarendon Press 1993).

come to be recognised—even by many physicalists—that the notion of supervenience is not strong enough to characterise physicalism.[2] For this reason, regardless of the other merits or inadequacies of supervenience, I will for the most part ignore the notion in this paper.

So the two major varieties of physicalism are identity theories and constitution theories. I prefer to classify physicalist theories in this way, rather than in terms of the more usual distinction between 'reductive' and 'non-reductive' physicalism, for two reasons. First, the concept of reduction is itself highly controversial: there are many accounts of what reduction actually is, and there is even a dispute among physicalists about whether there can be such a thing as non-reductive physicalism at all.[3] Second, it will emerge later in the paper that classifying physicalist theories in this way will divide those physicalist theories which have a problem with mental causation from those which do not. The standard view in the current literature is that it is 'non-reductive' physicalism which has this problem, and that it has it precisely because it is 'non-reductive'. I shall dispute this—the distinction between reductive and non-reductive physicalism does not help to show why the problem of mental causation arises.

3. The arguments for physicalism

Why believe in physicalism? The pioneers of the identity theory, U.T. Place and J.J.C. Smart, were chiefly concerned to show that the theory cannot be ruled out *a priori* for semantic or conceptual reasons.[4] Influential as it was, this argument is negative in character, and gives no positive reason for believing in physicalism. The only explicit positive argument given by Smart was based on Occam's Razor. But this would not impress Cartesian dualists, many of whom would argue that non-physical mental substances need to be posited in order to explain certain phenomena (e.g. conscious experience). Independent argument is needed.

We can identify three independent arguments for physicalism which have dominated the debate. The first is the argument advanced by David Lewis and D.M. Armstrong. As Lewis presents it, the argument has two premises, an *a priori* premise

---

[2]See (e.g.) Terence Horgan, 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World' *Mind* 102, 1993; Poland, *Physicalism* p.105; David Charles, 'Supervenience, Composition and Reduction' in D. Charles and K. Lennon, edd. *Reduction, Explanation and Realism* (Oxford, Clarendon Press 1992).

[3]See J.Kim 'The Myth of Non-Reductive Materialism' in *Supervenience and Mind* (Cambridge: CUP 1993); Brian Loar, 'Elimination versus Non-Reductive Physicalism' in *Reduction, Explanation and Realism*.

[4]See Place, 'Is Consciousness a Brain Process?' p.42, and Smart, 'Sensations and Brain Processes' p. 54 in C.V. Borst (ed.) *The Mind-Brain Identity Theory* (London: Macmillan 1970).

and an empirical one. The *a priori* premise is that mental properties are defined by their characteristic causal roles: their systematic patterns of relations to perceptions, actions and other mental states. (Or as Armstrong puts it, 'the concept of a mental state is the concept of a state of the person apt for bringing about a certain sort of behaviour'.[5]) Lewis defends this 'functionalist' view on the grounds that the idea of such a systematic pattern is implicit in commonsense psychological attributions of mental states.

Lewis's second, empirical, premise is that physical science is 'explanatorily adequate':

> there is some unified body of scientific theories of the sort we now accept, which together provide a true and exhaustive account of all physical phenomena. They are unified in that they are cumulative: the theory governing any physical phenomenon is explained by theories governing phenomena out of which that phenomenon is composed and by the way it is composed out of them. The same is true of the latter phenomena, and so on down to fundamental particles or fields governed by a few simple laws, more or less as conceived in present-day theoretical physics.[6]

(Lewis adds that this thesis is 'a traditional and definitive hypothesis of natural science—what scientists say nowadays to the contrary is defeatism or philosophy'.) The idea here is that any physical effect must be explicable in purely physical terms—in terms of purely physical phenomena. From this and the first premise it follows that mental properties are physical properties. For if the occupants of the mental causal roles were not physical, then given the second premise, they wouldn't be explicable in physical terms, and so wouldn't have have any physical effects. But since it is in their very nature to have effects they must be physical. (Lewis does not rule out nonphysical entities—it's just that they can have no physical effects, and so cannot be mental states.)

The second argument I shall call the *overdetermination* argument. It has been defended by Christopher Peacocke, James Hopkins and David Papineau, among others.[7] Unlike Lewis's argument, this argument deals with mental particulars (tokens) rather than mental properties (types). It can be expressed as follows. Suppose some token physical effect E has a token mental cause M, and that all physical effects have

---

[5] *A Materialist Theory of the Mind* (paperback edition, London: Routledge 1993). p.82. Armstrong's presentation of the argument is on pp.82-90.

[6] 'An Argument for the Identity Theory' *Philosophical Papers Volume I* (Oxford: Oxford University Press 1983) p.105.

[7] See Peacocke, *Holistic Explanation*(Oxford: Clarendon Press 1979); Hopkins 'Mental States, Natural Kinds and Psychophysical Laws' *Proceedings of the Aristotelian Society Supplementary Volume*, 1978, and D. Papineau, 'Why Supervenience?' *Analysis* 50, 1990.

complete physical causes. Then it follows that E has a complete token physical cause, call it P. So *either* M and P both cause E independently—that is, they overdetermine E—or M is identical with P. Since there is no overdetermination, M and P are identical.

The distinctive feature of this argument is the denial of overdetermination. Is this really so easy to deny? It seems only too easy to describe apparently possible cases in which overdetermination happens. Suppose a stabbing causes Sid's death, and simultaneously, a shooting causes his death. Both the stabbing and the shooting kill him, but either would have done so if the other hadn't. Is this sort of set-up really not possible?

Opinions differ over whether overdetermination cannot happen or just does not happen. Jaegwon Kim calls it 'absurd'. Christopher Peacocke is more cautious: he says the possibility of such overdetermination is something 'we ordinarily take to be false, and it is not clear why we should change the belief'.[8] Of course, there is a special problem about overdetermination for counterfactual analyses of causation, but if we put this issue to one side, the standard view among philosophers of mind seems to be that overdetermination cannot be a common feature of the actual world. As Stephen Schiffer puts it, 'it is hard to believe that God is such a bad engineer'.[9] I shall return to overdetermination in §9.

The final argument I shall consider is Davidson's famous argument in 'Mental Events', again for the identity of mental and physical events.[10] Davidson argues as follows: mental events interact with some physical events; when two events are related as cause and effect, they have descriptions under which they instantiate 'strict' laws of nature; and there are no strict psychophysical laws. These three claims are incompatible. As is well-known, Davidson reconciles the conflict by deriving the conclusion that all mental events are physical events. For if a mental and physical event interact causally, then they have descriptions under which they instantiate strict laws. But these laws must be physical, since there are no strict psychophysical laws. However, to fall under a physical law, the mental event must have a physical description, and if an event has a physical description it is a physical event. Therefore all mental events which interact with physical events are themselves physical events.

4. The 'completeness' of physics

---

[8]Kim, 'The Myth of Non-Reductive Materialism' p.281; Peacocke *Holistic Explanation* p.135.
[9]Schiffer, *Remnants of Meaning* (Cambridge Mass. 1987) p.148.
[10]'Mental Events' in *Essays on Actions and Events* (Oxford: Oxford University Press 1980).

All these arguments employ a common assumption about the causal structure of the physical world. This is most explicit in the second premise of Lewis's argument: the 'explanatory adequacy' of physics. What exactly does this mean? This depends partly on how we take the notion of explanation. If we think of explanations as being what scientific theories actually provide, then the premise says that physics can provide an explanation of all physical effects. This looks very implausible, even for a physicalist—most physicalists these days don't think that the doctrine requires that physics will explain everything. Contemporary physicalism allows that there are many kinds of indispensable but non-physical explanations of physical phenomena.

We can strengthen Lewis's argument by interpreting this premise in terms of causation. Understood this way, the premise says that all physical effects have physical causes which are adequate to determine (or fix the chance of) all other phenomena. The idea now is that fixing the physical causes (and physical laws) fixes all physical effects. This is the principle David Papineau calls 'the completeness of physics':

> All physical effects are determined or have their chances determined by prior physical [causes] according to physical law.[11]

I will follow Papineau in using the term 'completeness'. The idea behind this term is not that physics is a complete science of everything. The idea rather is that physical causes completely suffice for physical effects, or fix their chances: no other causes are *required* to bring about physical effects. That's the point of calling them 'complete'. (I shall from now on ignore indeterminism, since it is not directly relevant to the arguments I am discussing.)

The overdetermination argument also employs the completeness of physics as a premise. For presumably, the reason for believing that a particular effect has a complete physical cause is that all physical effects have complete physical causes. I shall take it that the premise that all physical effects have complete physical causes is substantially the same as Lewis's premise—the completeness of physics.

It is perhaps less obvious that the completeness of physics is a premise in Davidson's argument. But in fact it is a simple consequence of the nomological character of causation together with the anomalism of the mental. Take any physical effect (E) which is the product of a complete sufficient cause (C). The nomological character of causation says that C and E have descriptions under which they instantiate a strict law.

---

[11]Papineau *Philosophical Naturalism* (Oxford: Blackwell 1993) p.16. See also Horgan, 'From Supervenience to Superdupervenience' p.573.

But the anomalism of the mental says that this strict law cannot be a mental law of any kind—so it must be a physical law, and the sufficient cause C must be physical. Therefore any physical effect must have a sufficient physical cause—the completeness of physics again.

5. The general form of the arguments for physicalism

Having isolated this central assumption in all the arguments for physicalism, the general form of the arguments should now be obvious. It is this: to reconcile mental causation with the completeness of physics by identifying mental items with physical items. Lewis's argument does this by assigning mental states typical causal roles, and arguing that these causal roles are occupied by physical states. The overdetermination argument identifies mental tokens with physical tokens in order to block overdetermination by distinct causes—again, given the completeness of physics. And Davidson identifies mental and physical events to make mental causation consistent with the denial of psychophysical laws and the nomological character of causation, which together entail the completeness of physics.

There are obvious differences between the arguments, of course. Lewis's argument concerns types or properties, the other two concern tokens or particulars. And while Lewis, Armstrong and the defenders of the overdetermination argument consider their arguments to be *a posteriori*, Davidson's argument has a relatively *a priori* character. However, what I want to stress is the common structure in the arguments: to account for mental causation, given the completeness of physics, by identifying mental causes with physical causes. So it seems to me that John Searle gets the issue completely the wrong way around:

> one of the assumptions shared by so many traditional dualists and physicalists is that by granting the reality and causal efficacy of the mental we have to deny any identity relation between mental phenomena and the brain.[12]

On the contrary: it is because physicalists want to maintain the causal efficacy of the mental that they identify mental phenomena with phenomena in the brain. It seems, then, that if there are mental epiphenomena, there is no good reason—apart from the vague and inconclusive considerations surrounding Occam's razor—to say that they are physical.

---

[12]*Intentionality* (Cambridge University Press 1983) p.265

There is an important assumption hidden in these arguments. I call it the 'homogeneity' of mental and physical causation. That is, if the arguments are going to work, there must be a conflict between mental causation and the completeness of physics. But if this is so, the notion of causation is the same notion applied to the physical and the mental alike. So the general form of the arguments for physicalism is not that if there is mental causation, then it must be something so utterly weird and *sui generis* that we must, for reasons of parsimony and theoretical simplicity, reject it. For the obvious response to this is to acknowledge the *sui generis* notion of mental causation as just as weird (or not) as the *sui generis* notion of physical causation. There is no conflict—and thus no need for an identity thesis—if the notions of causation employed are so different. Another way of putting this point is that the arguments for physicalism must assume that the labels 'mental' and 'physical' as applied to causation are really transferred epithets—what is mental and physical are the relata of causation, not the causation itself.

6. Orthodox physicalism

The form of physicalism that is motivated by these arguments is (one or another version of) the identity theory. But it is notable that few contemporary physicalists actually accept identity theories. It is now generally considered that these theories are either too strong to be plausible, or too weak to be explanatory. Instead, physicalists tend to hold a version of the consitution theory mentioned in §2—the constitution theory has become the orthodox version of physicalism.[13] How does this shift from identity to constitution arise?

The most important cause of this shift is the influence of Putnam's variable realisation objection to the type-identity theory: it seems nomologically possible that many very different token physical entities could all be in the same type of mental state. So the type-identity theory is far too strong to be empirically plausible. But the token identity theory, on the other hand, seems too weak to be satisfactory—for what explains *why* these mental tokens are identical with these physical tokens? A solution to the mind-body problem is supposed to give an illuminating answer to the question of the relation between the mental and the physical. But it is hard to see how the token identity theory can do this. (An analogy might help make the objection vivid: it is hardly an explanation of why all US presidents have been white males to simply assert

[13]Versions of the constitution theory is defended and taken for granted by many of the contributors to two recent anthologies, *Mental Causation* edited by J. Heil and A. Mele (Oxford: Oxford University Press 1991); *Reduction, Explanation and Realism*, edited by D. Charles and K. Lennon .

the 'token identity' claim that each particular US president is identical with some particular white male.)

In the recent history of physicalism, this is where the notion of supervenience comes in: physical properties determine mental properties, but are not identical with them. However, supervenience is not strong enough to do the job of accounting for the mental-physical relation—so something else needs to be added to make supervenience a version of physicalism. This 'something else' is what I am calling 'constitution'. (Others use different terms: David Charles and Philip Pettit use 'composition', and in a recent survey of the supervenience issue, Terence Horgan optimistically uses the neologism 'superdupervenience'.[14])

This progression of theories—from identity theories *via* supervenience to constitution theories—is best thought of as an in-house debate among physicalists. It is rare that independent motivation is given for the constitution theory. The approach normally taken by physicalists is: 'we all know we have to be physicalists—the question is, what is an acceptable form of physicalism?'.[15] (The importance of this point will emerge at the end of this paper.)

In any case, it is undeniable that the orthodox version of physicalism is now the constitution theory: the theory that mental properties are constituted by physical properties. As I said in §2, I'm not going to discuss in detail what 'constitution' means, but part of the idea presumably is that the physical properties constitute the mental properties insofar as they are instantiated: my headache *now* is constituted by some physical properties instantiated in my brain. There is no commitment to the idea that uninstantiated mental universals—if there are such things—are constituted by uninstantiated physical universals. However, instantiated universals are still universals, and this means that the constitution theory is a 'token identity' theory only in the most anodyne sense. Suppose I instantiate the property *pain*, and I also instantiate a certain brain property, *B*, alleged to constitute the pain. There are then two complex entities: my instantiating the property *pain*, and my instantiating the property *B*. Though some philosophers (notably Kim) call such complex entities 'events', I shall follow Davidson and others in calling them 'facts' (they could also be called 'states of affairs').[16] This

---

[14]See Charles, 'Supervenience, Composition and Reduction'; Pettit, 'A definition of Physicalism'; Horgan, 'From Supervenience to Superdupervenience' p.566.

[15]See Poland, *Physicalism*, chapter 3.

[16]For Kim's view, see 'Causation, Nomic Subsumption and the Concept of Event' and 'Events as Property Exemplifications' in *Supervenience and Mind*. I follow Davidson, 'Events as Particulars' in *Essays on Actions and Events*, who in turn follows Ramsey, 'Facts and Propositions' in *Philosophical Papers* (Cambridge: Cambridge University Press 1990).

theory is a token identity theory only in the sense that the particular which has the property *pain* is the very same thing as the particular which has the brain property. That is: me. Since the properties *pain* and *B* are distinct, the facts which incorporate them must be distinct too. That I am a constituent of both these facts is certainly true—but it has very little bearing on the issue of physicalism, and it can only mislead to call this view a token identity theory.[17]

There is a position, however, which claims that it is the *instances* of the mental and physical properties that are identical, not the properties themselves. This view has been defended by Graham and Cynthia Macdonald.[18] The Macdonalds' idea is that an instance of a mental property (a 'property-instance') could be identical with an instance of a physical property without the properties themselves being identical. It might appear then, that this view is not the merely anodyne 'token identity' theory.

But what this position really amounts to depends on how we understand the notion of a property-instance. There are two fairly clear conceptions of instances of properties, but neither of these are what the Macdonalds want. The first conception is that an instance of a property is just the thing that has the property. Since I am tall, I am an instance of the property of *being tall*. But this is obviously not the Macdonalds' position. The second way of understanding property-instances is as *tropes*—the so-called 'abstract particulars' or 'particularised qualities' which some take to be the basic constituents of reality (for instance: *my tallness*). But the Macdonalds explicitly deny that their property-instances are tropes.[19]

What they actually say is that property-instances are 'events' (in something close to Kim's sense) whose 'constitutive components' are objects, properties and times.[20] (Notice how different this is from Davidson's conception of events, on which properties are not 'components' of events at all.) So property-instances are supposed to be entities distinct from the objects that have properties, and from the properties (universals) of which they are instances—even though these objects and properties are 'components' of property-instances.

It is now hard to see how property-instances differ from facts, as defined above: things having properties at times. The only difference seems to be this. It seems

---

[17]For a good statement of this point, see Papineau, *Philosophical Naturalism* p.24.

[18]See Cynthia and Graham Macdonald 'Mental Causes and the Explanation of Behaviour' *Philosophical Quarterly* 1986; 'Mental causation and non-reductive monism' *Analysis* 51 1991; Cynthia Macdonald, *Mind-Body Identity Theories* (London: Routledge 1989) chapters 4 and 5.

[19]See 'Mental Causation and Non-reductive Monism' pp.27-28. For tropes, see Keith Campbell, 'The Metaphysics of Abstract Particulars' *Midwest Studies in Philosophy VI: The Foundations of Analytic Philosophy* (Minneapolis: University of Minesota Press 1981).

[20]See Kim, 'Causation, Nomic Subsumption and the Concept of Event' .

plausible to say that the fact that *I am in pain at t* and the fact that *I am in brain state B at t* are the same fact just in case *being in pain* is the same property as *B*. But this is not true of the Macdonalds' property-instances. Their view is that the one instance contains as 'components' the property *being in pain* and the property *B*. This is what they mean by saying that mental and physical properties can be 'instantiated in a single instance'.[21] Though it is true to say that mental and physical property-instances are *identical*, it could be rather misleading—what the Macdonalds mean is that a single property-instance has as 'components' a mental property and a physical property.

However, it is not obvious why someone who holds the fact theory *cannot* hold that a fact can contain as 'components' both a mental property and a physical property. And if so, then there is not much to choose between the fact theory and the property-instance theory, and the Macdonalds' view is just the fact theory in different terminology.[22] Progress in this area is frustrated by the lack of an adequate theory of facts, and of what it is for a property to be a 'component' or 'constituent' of a fact—until we know this, we don't know whether it is true that a fact cannot have a mental and a physical property as components.

For this reason, I shall treat the Macdonalds' theory as a version of the constitution theory. What is doing the work in the theory is the idea that mental and physical properties are (in some way) 'united' but not identical. The fact that they are united in one complex entity does not differentiate thir theory sufficiently from the constitution theory—according to which two facts could be seen as 'united' (in some way) in one complex entitity. And as we shall see, the property-instance theory faces the same problem that the constitution theory faces.

The move from identity to constitution (and related notions) is sometimes seen as a fairly innocuous one, of little ontological consequence. In a recent attempt to define physicalism, for example, Philip Pettit treats identity as merely a special case of 'composition'.[23] But it seems to me that the move is very significant, since once it is accepted, physicalists have to face the problem of mental causation.

7. The problem of mental causation

The story so far: the arguments for the various identity theories attempt to reconcile mental causation with the completeness of physics. However, the most common form

---

[21]'Mental causation and non-reductive monism' p.28.

[22]See John Heil,*The Nature of True Minds* (Cambridge: Cambridge University Press 1992), pp. 135-139, who defends a view very similiar to the Macdonalds's view and to the orthodox version of physicalism. He calls it 'realisation'.

[23]'A Definition of Physicalism' *Analysis* 1993

of physicalism is not the identity theory, but one or another version of the constitution theory. I now want to move on to what many orthodox physicalists consider to be one of their main problems: the problem of mental causation.

We can construct a simple version of the problem as follows. Causes have their effects in virtue of their properties. If I throw a brick at the window and the window breaks, this will be because of certain properties of the brick. (It doesn't matter here if these are taken as properties of the brick, or of an event—the event of the brick hitting the window.) But not all properties of a cause are responsible for its effects. It is not the colour of the brick, or its sheen, or its 'relational properties' like its being made in Walthamstow, or its being thrown at exactly 4.05 pm, that are responsible for the window's breaking.

Now suppose that particular mental states are not identical with particular physical states. We can ask, are the mental properties of causes responsible for their effects? This raises a dilemma: if the mental properties are responsible for the effects, then either the completeness of physics is false or the effects are overdetermined. And neither of these options are acceptable to physicalists. On the other hand, if the mental properties of the cause are not responsible for its effects, then epiphenomenalism is true: the mental makes no causal difference. So orthodox physicalism seems either inconsistent or epiphenomenalist—this is the problem of mental causation for physicalists.

Notice that this argument assumes what I call in §5 the 'homogeneity' of mental and physical causation. If homogeneity doesn't hold, then there is no problem. Notice too that because of the way the problem is posed—'do causes have their effects in virtue of their mental properties?'—the problem obviously cannot arise for type identity theorists.

As a number of writers have observed, there are actually two separate questions about epiphenomenalism involved in this problem.[24] The first is whether any particular mental states/events are causes at all. The second is whether mental states/events are causes in virtue of their mental properties. It might seem possible to hold that mental states/events are causes, but that they are not causes in virtue of their mental properties—if for example, token mental states/events were identical with

[24]See Ernest Sosa, 'Mind-Body Interaction and Supervenient Causation' *Midwest Studies in Philosophy* 9 (1984), p.278; Brian McLaughlin, 'Type Epiphenomenalism, Type Dualism and the Causal Priority of the Physical', in *Philosophical Perspectives 3: Philosophy of Mind and Action Theory*, edited by James E. Tomberlin, (Atascadero: Ridgeview 1989), and Yablo, 'Mental Causation' pp.248-250. As Yablo notes, the distinction derives from C.D. Broad: *The Mind and it Place in Nature* (London: Routledge & Kegan Paul 1925) p.473.

token physical states/events. I shall look at positions like this in the next section. But it is worth pointing out at this stage that if we accept the principle that causes have their effects only in virtue of their properties, then the distinction between the two questions effectively collapses.

A distinction we must now make, however, is between our problem of mental causation and the problem of the 'causal efficacy of content'.[25] The problem about the efficacy of content is the problem of how the contents of intentional states can be relevant to the effects of those states. But the problem of mental causation arises for all mental states, not just for those with intentional contents. Furthermore, the problem about content is clearly not that it is *mental*—indeed, this idea scarcely makes sense. For suppose, for the sake of argument, that everything is physical, and that a state's possession of intentional content has been reduced to purely physical facts about that state and its relations to the environment. Then the question can still arise: is it in virtue of having this content that an intentional state has its effects? The fact that this question can arise shows that even if a type identity theory were true, the problem of the efficacy of content remains. But as I have construed the problem of mental causation, the problem couldn't arise if type identity theories were true.

Nor is the issue about whether there are any adequate mental *explanations*—explanations which cite mental states or events as explanatory of behaviour. If this were the issue, then there could hardly be a problem at all—for it is surely uncontroversial that mental concepts are used successfully to characterise and explain behaviour. The question is what makes these explanations work—specifically, do they work because mental states/events are among the causes of behaviour?

It is for this reason (among others) that a mere counterfactual criterion of mental efficacy will not solve the problem. It will not do to say that mental efficacy is ensured by the truth of counterfactuals like 'if the mental property had not been instantiated, then the physical effect would not have occurred'. For it is well-known that such counterfactuals could be true without the antecedent picking out a cause of the phenomenon in question. For example, the truth of these counterfactuals is consistent with the mental property and the physical effect being effects of a common cause.[26]

---

[25]For a clear account of the problem of the efficacy of content, see Ned Block, (1990) 'Can the Mind Change the World?', in *Meaning and Method*, edited by George Boolos (Cambridge: Cambridge University Press 1990). Some writers who discuss the efficacy of content are best interpreted as tackling the mental causation problem too: see for instance Gabriel Segal and Elliott Sober 'The Causal Efficacy of Content' *Philosophical Studies* 63 1991, and Michael Tye *The Imagery Debate* (Cambridge Mass.: MIT Press 1992) chapter 8.

[26]For more on counterfactual criteria of 'causal relevance', see Ernest Le Pore and Barry Loewer, 'Mind Matters', *Journal of Philosophy* 84, 1987; Jerry Fodor, 'Making Mind Matter More' in *A Theory of*

Although I distinguish here between causation and explanation, I do not need to say much about what the debate assumes about the nature of causation, since most of the debate does not depend on any specific theory of causation. (An important exception is Davidson's theory, which I shall discuss in the next section.) All participants can assume for example, that deterministic causes are necessary and sufficient (in the circumstances) for their effects, or that causes must raise the chances of their effects—make those effects more probable than they would otherwise have been—or that their existence implies the existence of a law under which they fall.

There is one assumption about causation which is essential to the argument. This is that causes have their effects in virtue of some of their properties. In characterising causes we pick out certain features of them—'causally efficacious properties'—in virtue of which they have the effects they do. This is an assumption I need not defend here, since it is taken for granted by almost all participants in the debate.

But who are the participants in this debate? And who exactly should be worried by the problem?

8. Whose problem is this?

The contemporary mental causation debate arose chiefly through criticism of Davidson's anomalous monism.[27] Davidson's critics in effect posed our question—'do mental events have their effects in virtue of their mental properties?'—and argued that if Davidson says that mental events do have their effects in virtue of their mental properties, then there must be psychophysical laws; but if he says that mental events have their effects in virtue of their physical properties, then then the mentality of mental events is causally redundant. So anomalous monism is either inconsistent or epiphenomenalist.

But despite the extensive discussion of this point in the literature, this is simply not a problem for Davidson at all. This is because for Davidson, causation is a relation between particular events regardless of which properties they have—or to put it in Davidson's nominalistic way, regardless of how they are described. As Davidson himself says,

---

Content and Other Essays (Cambridge, Mass.MIT Press 1990); Gabriel Segal and Elliott Sober 'The Causal Efficacy of Content'; and Brian Leiter and Alexander Miller 'Mind Doesn't Matter Yet' *Australasian Journal of Philosophy* 72, 1994.

[27]See Ted Honderich, 'The Argument for Anomalous Monism' *Analysis* 1982; also Ernest Sosa, 'Mind-Body Interaction and Supervenient Causation'.

> it is events that have causes and effects. Given this extensionalist view of causal relations, it makes no literal sense ... to speak of an event causing something as mental, or by virtue of its mental properties, or as described in one way or another.[28]

This is not epiphenomenalism, by Davidson's lights—it would be epiphenomenalism if it made sense to say that the physical features of mental events were 'more efficacious' than the mental. But it does not: for it is also 'irrelevant to the causal efficacy of *physical* events that they can be described in the physical vocabulary'.[29]

The central point here is that Davidson rejects the principle mentioned in the last section that causes have their effects in virtue of some of their properties. He rejects this not simply because of his rejection of properties, but because he holds that causation is a relation between particulars. We could put the point this way. If some properties of a cause make it 'more efficacious' than others, then some ways of describing the cause are better than others—some ways pick out the causally relevant features, and others don't. But on Davidson's theory, there is no way of describing a cause which describes it as 'more efficacious' than other ways. All true descriptions of the cause describe something efficacious—otherwise they would not be true descriptions of the cause.

This is not to say that Davidson cannot give any answer to the question, 'was it in virtue of the brick's mass or its colour that it broke the window?'. But he will regard this question not as a question about the efficacious *properties* of the cause, but as a request for an informative causal *explanation*. And as we saw in the last section, the issue of explanation is largely irrelevant to our problem.

This point is so obvious that it prompts the question: why do so many people think Davidson does face the problem of mental causation? Part of the reason, I suspect, is that sometimes the idea of a token identity theory is defined in the philosophical literature independently of Davidson's theory of causation.[30] Kim's conception of 'events' as incorporating properties is also well-known and widely accepted, as is the link between properties and causation; if these ideas are in the background to a reading of 'Mental Events', then it seems very natural to raise the problem of mental causation for Davidson. It's as if people think: 'Davidson has shown how to be a physicalist without being a reductionist. But now we have to decide which theory of events and which theory of causation to adopt'. But if you accept Davidson's

---

[28]'Thinking causes' in Heil and Mele (edd.) *Mental Causation*, p.13

[29]'Thinking causes' p.12

[30]See, for example, Horgan, 'From Supervenience to Superdupervenience' p.563.

argument for his conclusion, then it cannot be a further question whether you accept his theory of events and causation.

More importantly, though, is the fact that there is something very unsatisfactory about Davidson's (implicit) denial of the thesis that causes have their effects in virtue of some of their properties. The point of the thesis is to mark a crucial distinction between those properties whose instantiation genuinely makes an objective causal difference (in a particular case) and those whose instantiation does not. As I said above, it is not that Davidson cannot mark this distinction. But he will mark it as a distinction among explanations, not among properties of things. The trouble is that Davidson's theory seems to leave us unable to answer the question of why certain explanations are better than others by invoking the efficacious *features* of reality. So my view is that although Davidson is not troubled by the problem of mental causation, major difficulties with his theory lie elsewhere.

The fact that Davidson does not have a problem of mental causation is the reason why I do not want to characterise those who face the mental causation problem as 'non-reductive physicalists'—since Davidson is the pre-eminent non-reductive physicalist, yet he does not have to face the problem.

But what about the 'property-instance' view—does it have to face the problem? Its defenders are keen to emphasise Davidson's distinction between the relation of causation, which can be characterised in a purely extensional language, and causal explanations, which are non-extensional.[31] But given their conception of events as instances of properties, it is not obvious that they can see things in Davidson's way. For Davidson's events are basic particulars, the values of variables bound by first-order quantifiers. It is easy to see how the distinction between explanation and causation applies to this theory, but it is obscure how the distinction applies to Kim's conception of events or property-instances. (It is no accident that those, like D.H. Mellor, who reject Davidson's conception of the relata of causation also reject his thesis that causation can be characterised in a purely extensional language.[32])

The natural response for the property-instance view is to turn itself into a version of the constitution theory—as I claimed it does in the last section. But then it too will face the problem of mental causation which is endemic to that view. Take a mental fact or event, which seems to be the cause of a certain physical fact or event. If the completeness of physics is true, and there is no massive overdetermination, and

---

[31]Cynthia and Graham Macdonald, 'Mental Causation and Non-Reductive Monism' p.24.

[32]See D.H. Mellor, 'The Singularly Affecting Facts of Causation' in *Matters of Metaphysics* (Cambridge 1991), pp.211-213.

causation depends on properties of things, then the problem is plain: the mental fact cannot be the cause.[33]

This is why it is the consitution theory, however it is described—in terms of supervenience, composition, realisation or dependence—which has the problem of mental causation. So Jerry Fodor could hardly be more wrong when he says that 'mind-brain supervenience is the best idea anyone has ever had about how mental causation is possible'.[34] Apart from in the limiting case of type identity—where supervenience holds trivially—the supervenience/constitution/realisation theory gives us no idea whatsoever about how mental causation is possible.

9. The assumptions behind the problem

On the face of it then, orthodox physicalism is untenable because it has no way of reconciling mental causation with the completeness of physics. Of course, there have been many attempts to make this reconciliation. But it turns out that all these attempts end up denying one of the assumptions that motivate physicalism in the first place.

To see why, we have to spell out the assumptions behind the problem of mental causation, as formulated above. They are the following:

(A) Causes have their effects in virtue of some of their properties.

(B) There is mental causation.

(C) The completeness of physics is true.

(D) There is no overdetermination.

(E) Mental and physical causation are 'homogeneous'.

It seems to me that the only real way to reconcile all these assumptions is to be a type identity theorist: to identify mental properties with physical properties. (In fact, this ought to be obvious because of the similarity between the the general form of arguments for identity theories and the assumptions (A)-(E).)

But if type identity theory is rejected, as it is by most parties in the debate, then the problem can only be solved by rejecting one or more of the assumptions (A)-(E). As we saw in the last section, Davidson's identity theory rejects (A). What about the other assumptions? An epiphenomenalist will reject (B). But surely this is the last assumption we should reject—that our minds make our bodies move is not a piece of philosophical theory, but something which theory should explain.

---

[33]For further criticism, see Yablo, 'Mental Causation' p. 259 n32, and E.J. Lowe, 'The Causal Autonomy of the Mental' *Mind* 102, 1993, p.631.

[34]Fodor, *Psychosemantics* (Cambridge, Mass.: MIT Press 1987) p.30. Schiffer makes this same criticism in *Remnants of Meaning* p.154.

Contemporary philosophers rarely consider the possibility of denying (C), the completeness of physics. One reason for this seems to be that they think that denying completeness commits you to Cartesian dualism.[35] But this seems to me to be a mistake. For it ignores the availability of a position according to which physical effects can have many kinds of cause, none of which have the features Descartes attributed to the mind, and not all of which are physical (in the precise sense of 'physical' which physicalism needs).

For simplicity, let us ignore indeterminism here. Then what the completeness of physics says is that every physical effect is determined by purely physical causes in accordance with physical laws. Or in other words, purely physical causes suffice for the occurrence of every physical effect. But whatever else 'suffice' means, it must at least mean 'suffice in the circumstances'—even when we ignore determinism. Striking a match suffices for the match to light only given the presence of oxygen, imflammable material, and so on. Similarly, the presence of oxygen suffices for the match to light in the presence of the striking and the imflammable material—so it too is a cause of the lighting of the match.

So likewise, suppose someone throws a brick at a window because they want to, and it breaks. I can agree with physicalists that in the circumstances, the person's brain state suffices for their muscles to move, and for the brick to fly through the air, and for it eventually to break the window. But if the circumstances also include the person's beliefs and desires, then these too will suffice, given the other circumstances, for the window's breaking. So in any plausible sense in which physical causes suffice for their effects, mental causes can do so too.

But does that mean that there are 'gaps' in the chain of physical causation which the beliefs and desires must 'fill'? Why should there be? And what does it even mean to suppose that there must be these 'gaps'? As Tyler Burge says, thinking in these terms is 'thinking of mental causes on a physical model—as providing an extra "bump" on the effect'.[36] And the tendency of many physicalists to talk in terms of causes requiring 'forces' or 'oomph' shows that they are already thinking in terms of a maodel of causation which makes autonomous mental causation hard to understand.[37] This models of causation can be deeply misleading—bumping and 'impacting' are just a

[35]See e.g. David Papineau, 'The Reason Why: Response to Crane' *Analysis* 1991.
[36]Burge, 'Mind-Body Causation and Explanatory Practice' in Heil & Mele (edd.) *Mental Causation* p.115. See also E.J. Lowe, 'The Causal Autonomy of the mental' p.637.
[37]See Pettit, 'A Definition of Physicalism' p.219; Horgan, 'From Supervenience to Superdupervenience' p.572.

special cases of causation, not notions which illuminate causation. Likewise with the idea that 'gaps' need to be filled.

To reject the completeness of physics, then, is not to reject the claim that physical causes are sufficient causes of all physical effects. It is to reject the claim that *only* physical causes are sufficient causes of physical effects. Therefore, to hold that physical causes are sufficient causes of physical effects is not to hold that if the mental causes had not been there, then the physical effects would still have happened. And this is just as well, since prior to accepting any more extreme physicalist hypothesis, we have no good reason for believing it is true.[38]

According to the position I am describing, physical effects, like any effects, can have many causes—some of them are mental, and some of them physical. By itself, this innocuous claim is not in conflict with the laws of physics—e.g. the conservation laws—or with perfectly legitimate methodological principles—e.g. the explanatory requirement to look for underlying mechanisms for phenomena. While these principles might form part of the philosophical motivation for the completeness of physics, it is important to keep completeness separate from these other principles if non-physicalism is to avoid guilt by association.

Now just as Cartesian dualism does not follow from the denial of the completeness of physics, so physicalism does not follow from affirming it. For (D) can be rejected: actions are overdetermined by mental and physical causes. As we saw in §3, overdetermination is not normally ruled out on *a priori* grounds. It becomes hard to believe when it is supposed to be happening all the time, because this would involve massive coincidence. However, if we believe that mental and physical states are linked by psychophysical laws—a claim which is defensible on independent grounds—then overdetermination would not be a coincidence: it would be a matter of natural law that the mental and the physical causes both bring about the effect. Suppose I have a pain which causes me to go to the cupboard and get an asprin. The pain and the brain state are both actual causes of my going to the cupboard, and since they are linked by a psychophysical law, it is not a coincidence that both causes result in me going to the cupboard.

Notice too that this view can also remove the counter-intuitive consequence that I would have still gone to the cupboard even if I hadn't had the pain. For it isn't true that if I hadn't had the pain I would still have gone to the cupboard. The reason is that because the pain and the brain state are linked by a psychophysical law, the closest

[38]For an excellent discussion, see Burge, 'Mind-Body Causation and Explanatory Practice' esp. pp.111-117.

worlds in which I didn't have the pain, are worlds in which I didn't have the brain state either.[39]

These are two non-physicalist ways to escape the problem of mental causation. However, as plausible as these unpopular options are, the standard line is to reject (E), the homogeneity of mental and physical causation. That is, if there are mental causes, they are not causes 'in the same way' that physical causes are. Some philosophers are more explicit about their rejection of (E) than others are. Most explicit are certain physicalist theories of the efficacy of intentional content. Dretske, for instance, explicitly distinguishes between the way in which intentional states are 'structuring' causes of behaviour, and the way in which physical states are 'triggering' causes. And Frank Jackson and Philip Pettit distinguish between the sorts of explanations which we give of intentional phenomena—'program explanations'—and ordinary causal explanations, which they call 'process explanations'.[40] But since the efficacy of content is not the topic of this paper, I shall ignore these theories here.

One influential orthodox physicalist account of mental causation is Kim's theory that mental causation is 'supervenient causation'. Generally, X superveniently causes Y iff X supervenes on X*, and Y supervenes on Y* and X* causes Y*. In the mental case, 'a mental event M causes a physical event P... because M is supervenient upon a physical event P*, and P* causes P'. [41] Whatever the merits of this idea, it is plain that if completeness and the denial of overdetermination are retained, then whatever supervenient causation is, it cannot be the same relation as the subvening physical causation. This is why Kim's position involves a denial of homogeneity.[42]

Another attempt to reconcile the assumptions (A)-(E) is the claim made by Block and Papineau that mental properties are second-order properties: a mental property M is the property of having some property P the instantiation of which has certain causes and effects.[43] If it is insisted, against the charge of epiphenomenalism, that M is nonetheless 'responsible' for these effects, then we must ask: in what sense? If overdetermination and the completeness of physics are denied, then it cannot be the

---

[39]The point is Hugh Mellor's: see Tim Crane & D.H. Mellor, 'Postscript' in Moser & Trout (edd.) *Materialism*.

[40]See Dretske, *Explaining Behaviour* (Cambridge Mass.: MIT Press 1988); Jackson and Pettit, 'Functionalism and Broad Content' *Mind* 97, 1988.

[41]See Kim 'Epiphenomenal and Supervenient Causation' in *Supervenience and Mind*, p.106. The supervenience in question is 'strong' supervenience. See also Segal and Sober, 'The Causal Efficacy of Content' and Tye, *The Imagery Debate* chapter 8, esp.pp.147-8.

[42]Actually, Kim seems to waver between denying (B) and denying (E)—this is because he thinks supervenient causation is 'epiphenomenal causation': see 'Epiphenomenal and Supervenient Causation'

[43]'Can the Mind Change the World?' pp.155-166; *Philosophical Naturalism* p.25.

same sense in which P is responsible for the effects in question. So homogeneity is denied.

The same general pattern applies to the ingenious recent suggestion of Stephen Yablo's, that mental properties are determinables (in W.E. Johnson's sense) of which their physical realisers are determinates. For example, just as being coloured is a determinable which has many determinates (green, red etc.) so being in pain is a determinable which has many determinates—the types of brain properties which are the realisers of pain. The solution to the problem of mental causation then relies on the idea that just as there is no overdetermination of the effects of an object's being green by the fact that green objects are also coloured objects, so there is no risk of overdetermination of the effects of physical properties by the mental properties which they realise. Determinates and their determinables are not 'causal rivals'.[44]

Yablo denies that mental and physical facts are identical. But since he also thinks this is compatible with the completeness of physics—'everything that happens is in strict causal consequence of its physical antecendents'[45]—then he has to reject homogeneity. He does this by distinguishing between 'causal sufficiency' and 'causal relevance'. A mental property can be causally relevant to the effects of its instances, but only the physical properties are causally sufficient.

It is sometimes objected that these various responses to the problem of mental causation are really just thinly disguised forms of epiphenomenalism. But without going into the details of the theories of causation which the responses presuppose, it is hard to adjudicate this question. Moreover, orthodox physicalists tend to respond that (for example) supervenient causes are still causes 'in some sense', and that it is only if we have a crude conception of causation that we think mental causes ought to be causes in 'the same sense' as physical causes.[46] For these reasons, there is a tendency for the debate to descend into a fruitless exchange of intuitions about what exactly 'epiphenomenon' means. So I want to try a different approach and look at the effect that denying homogeneity has on the motivation for orthodox physicalism.

Orthodox physicalism, as we have seen, gives rise to the very problem physicalism was originally introduced to solve. In order to solve this problem in orthodox physicalist terms, the typical manoeuvre is to deny the homogeneity of mental and physical causation. But the problem with denying homogeneity is that it is now impossible even to *state* the orginal motivation for physicalism: the conflict

---

[44]Yablo, 'Mental Causation' p.259.
[45]Yablo, 'Mental Causation' p.279.
[46]See Dennett, 'Real Patterns' *Journal of Philosophy* 1989 pp.

between mental causation and the completeness of physics. So there is no longer an original motivation for any kind of physicalism—there is no good reason for saying that these mental phenomena are 'constituted by' or 'realised by' physical phenomena. In effect, the point is implicit in Lewis's 1966 paper: it's only insofar as mental states have effects *in the very same sense that physical states have effects* that we need to think of them as physical states. As we have seen, orthodox physicalism denies that mental states have effects 'in the very same sense' that physical states have effects. For it now says that there are mental phenomena which are causes in their own way; and there are physical phenomena which are causes in a different way. Whether or not such distinctions between different kinds of causation are ultimately tenable, it seems plain that little is added by saying that these mental phenomena are 'ultimately physical'. And it seems an empty terminological decision to call the resulting position 'physicalism'—except perhaps to put on the record one's differences with Descartes. But if this is all physicalism means, then there is no need to bother with formulating any more precise version of the doctrine.

10. Conclusion

So what is the lesson of the recent mental causation debate? I am not denying that someone can come up with a philosophical account of the relation between mind and body—along the lines of constitution/realisation/supervenience—and call it 'physicalism'. My point about these varieties of physicalism is rather that they have no clear philosophical motivation. The lesson of the mental causation debate, then, is that there is no well-motivated physicalist position which is not an identity theory.

This conclusion resolves our original puzzle. But a further puzzle remains. Recent philosophy of mind has been dominated by attempts to describe and defend adequate versions of 'non-reductive' physicalism. Often these attempts are responses to the mental causation problem, but they are also presented as being of independent philosophical interest. What is clear in this debate is that mental causation is a problem for those who reject the identity theory. However, given the standard response to this problem, it is puzzling that physicalists think there is still a need to answer the question: 'what is an adequate version of physicalism?'. If this is a significant question, then physicalists still have to explain why.[47]

[47]I am very grateful to Doug Ehring, Keith Hossack, Mike Martin, Lucy O'Brien, David Papineau, Sarah Patterson, Gabriel Segal, Barry Smith, Scott Sturgeon and Bernhard Weiss for many discussions which have helped me. Earlier versions of some of this material have been presented to audiences at the Universities of Birmingham, Leeds, Liverpool, Manchester, Reading and Wales (Lampeter). I am grateful to participants in those discussions for their reactions and comments.

Department of Philosophy
University College London
Gower Street
London WC1 6BT
tim.crane@ucl.ac.uk