**Physicalism without Causal Closure**

**David Papineau**

New York/New Jersey Philosophy of Science Group

Monday 2 March 2015—4pm-6pm Room B01 Goddard Hall NYU

A The Causal Argument for Physicalism

1. Why be a physicalist? Plenty of philosophers are not ‘antecedent physicalists’.

2. What is ‘physicalism’? Let’s not get bogged down. If we are interested in the status of life and mind, we may as well equate ‘physical’ with ‘(metaphysically fixed by) things that can be identified non-organically’, ie by terms referring to things found outside actual living and/or sentient bodies. (Re Hempel’s dilemma, note we aren’t looking to physics to define the ‘physical’ realm—though it may of course tell us about its nature.)

3. Premise 1 Mental states have physical effects

Premise 2 All physical effects have full physical causes (to extent they have full causes at all)

Premise 3 Not all physical effects of mental causes are metaphysically overdetermined

Conclusion The mental states (in 1) must be same as (part of) the physical causes (in 2).

4. Note how premises 1 and 3 knock out epiphenomenalism and ‘overdeterminationism’ respectively, and are effectively appeals to simplicity. (Note also how so understood premise 3 really only carries weight against ‘strong’ not ‘weak’ overdeterminationism.) But if we are appealing to simplicity, why do we need causal closure? Answer—to knock out interactionist dualism. Those who think that all we need is the ‘simplest explanation of the mind-brain correlations’ are forgetting that interactionism is itself a perfectly simple view, and that until it’s eliminated there no reason to believe in pervasive synchronic mind-brain correlations.

5. (Newtonian) interactionist dualism, with ‘vital’ and ‘nervous’ forces, was orthodoxy until relatively recently, long after the acceptance of the conservation of energy. I’d say it really died only mid-20thC (cf Huxley and Hodgkin).

B Causation is Macroscopic

6. OK. Today I’m worried about the way the argument relies on the notion of ‘cause’. Let me come at this issue sideways. Consider the ‘exclusion argument’ against non-reductive physicalism. The standard set-up: some mental state M supervenes on physical state P without being reducible to it; some further physical effect P\* apparently results from M; so how now can M be the cause of P\*? After all, M is a different property from P, so P already causes P\*, M seems otiose. Some eg Kim and Lewis agree—it’s the specific neural state P that causes your behaviour, not any more generic M. Others eg Bennett and Loewer argue that both P and M causes P\*—the resulting systematic overdetermination is benign, unlike systematic overdetermination by metaphysically independent causes. But I like the idea that M and not P causes P\*. (So boo to Kim: ‘The question is not whether P should be considered a cause of P\*; on anyone’s account, it should be’ 1993.)

7. This idea goes back to Stephen Yablo in the 1990s. The bull can be angered by the cape’s redness even if the cape is also more specifically crimson. The cause is the ‘proportional’ fact. Peter Menzies and Christian List have more recently argued that Woodward’s difference-making account of causation supports the same conclusion that it is M that causes P\* not P. Many respond that that Yablo and Menzies might be right about anthropocentric explanation, but surely ontological causation is to be found at the level of maximally specific physical problems. No doubt they take it to be obvious that P is a cause of P\* because they assumes that causation is constituted by the way maximally precise physical arrangements evolve in accord with basic dynamical laws.

8. However causation is asymmetric in time, but basic dynamics is not. So causation must involve something more than basic dynamics. If you are given a basic description of the temporal evolution of a physical system, plus the laws of dynamics, but not told which temporal direction is which, you won’t be able to read this off from the description. But if you can discern the causes and effects within a physical process, then this alone will tell you which way time is directed. Causes always come earlier than their effects, and so a specification of causal structure will tell you which temporal direction is which.

9. Of course, this would be trivial if the difference between ‘causes’ and ‘effects’ simply consisted in the former occurring earlier than the latter. But the difference between causes and effects lies deeper than this, and can be discerned prior to any independently given information about temporal order. Causal relations have a characteristic probabilistic signature which is asymmetric in time. It is this signature that the recent tradition of ‘Bayesian nets’ exploits in order to infer causal structure from probabilistic correlations. This argues that there must be something in the nature of causation that orientates it in time, and goes beyond the temporally symmetric structures of basic dynamics.

10. The second law of thermodynamics is also inexplicable in terms of basic dynamics alone, given that it refers to a specific direction in time. So an explanation of the second law needs to invoke assumptions that go beyond basic dynamics: first, that entropy was low in the past, and, second, that there is a certain probability distribution over all the precise physical microstates consistent with given ‘macrostates’ of temperature, energy, entropy, and so on. (See Albert Time and Chance 2000). I take causal asymmetry to have an analogous basis. The asymmetric correlational structures displayed by causal relationships suggest that causal processes too are essentially macroscopic, constituted by the nature of past facts together with probability distributions over the maximally specific microstates that can realize given macrostates.

11. Are thermodynamic processes, and causal relationships along with them, really essentially macroscopic? If thermodynamic processes like entropy increase get explained in terms of particle physics (plus other assumptions) then why count them as macroscopic? Take a volume of gas that is hot in one half and cold in another. Statistical mechanics tells us that in a while the temperature will almost certainly be uniform throughout. Now, you could in principle have analysed this particular system by applying basic dynamics to the precise initial conditions. But this microanalysis would owe nothing to the general principle that almost any system in that initial macrostate would end up at a uniform temperature. To bring out the general principle, you need to ‘throw away’ the information about the precise microstate, and note instead that the system is in a macrostate which is overwhelmingly likely to (be realized by a microstate that will) end up with a uniform temperature. Similarly, I suggest, with causation. If you focus on the precise microstate of some physical process, you will lose sight of causation. The causal structure of the world depends on probabilistic facts about the ways in which given macrostates are realized at the micro-level, rather than on the actual micro-realizations themselves.

12. Consider the Bayesian net strategy once more. Many philosophers nowadays start with the DAGs. But the interesting thing is the way that these can be derived from undirected correlations between the variables at issue. This derivation of course must hinge on some assumptions about the relationship between causal and correlational facts. Philosophers in this area are surprisingly uncurious about the connection between these assumptions and the nature of causation. When we do investigate this, we find support for the thermodynamic-style account of causation.

13. A salient feature of the cause-correlation connection is the common cause structure: the joint effects A, B of a common cause C (but not the joint causes of a common effect) will display an unconditional correlation that is ‘screened off ‘ when we control for the common cause. Why so? Well, suppose ‘pseudo-indeterminism’: CX v Z <-> A and CY v Q <-> B (C seems a probabilistic cause of A, and B, only because we ignore X, and Y). Then it is easy to show that you will get a screened-off correlation between A and B iff the ‘error terms’ X, Y, Z, Q are probabilistically independent.

14. Dan Hausman has codifed the Bayes Net assumptions like this: A causes B if and only if A and B are correlated, and everything correlated with A is correlated with B, and something correlated with B isn't correlated with A (Causal Asymmetry 1998). The idea here is that we can distinguish the effects from the causes by noting that they have probabilistically independent sources of variation (cf the X, Y, Z, Qs above). The issues are complicated, but I take points like these to argue that asymmetric causal structure is constituted by probabilistic distributions over the ways in which macroscopic causes are realized.

15. We don’t intuitively think of causation in statistical terms, but in terms of things bumping into each other. Thus Sosa and Tooley

“. . . it is not a necessary truth that any world containing causally related events is one where events typically have more effects than causes. The world . . . could have been a very simple one, where there were no causal forks . . . Lewis’s analysis cannot be sound, therefore, since there are logically possible causal worlds for which it yields the wrong results . . .” (Sosa and Tooley Causation 1993).

Well, of course there are conceivable scenarios which contain causal relations but lack asymmetric probabilistic structures. And if I were defending my view of causation as a piece of conceptual analysis, then such scenarios would suffice to refute it. But I’m not doing conceptual analysis, but proposing a synthetic metaphysical theory, which aims to uncover the nature of causation a posteriori. You can’t argue against theories of this kind by appealing to merely conceivable scenarios, any more than you can argue against orthodox chemistry by appealing to the conceivability of a world with water but no H20. If worlds with causation but no probabilistic structure were metaphysically possible, then that would be a problem. But if causation is indeed constituted by temporally asymmetric probabilistic structure, then there is no metaphysical possibility of the one without the other, however much this may be conceivable

16. OK. Back to Yablo-Menzies-List. We now have some serious theoretic support for their view that the higher-level M can be the cause and not its more specific realizer P.

17. Note that this doesn’t mean, fortunately, that the more specific neuronal P can never cause anything. It will come out differently if the effect P\* is not your arm moving as such, but your arm moving in the precise way that it did on this occasion. Now M is not specific enough for the result, and it is the specific neuronal arrangement P that is just right. Didn’t I say that causation is essentially macroscopic? However, this idea isn’t inconsistent with some neuronal P being a cause. The reason is that P needn’t constitute a maximally specific physical state, and so still leaves room for the kinds of probabilistic facts which I say are essential to asymmetrical causal relationships. It is only at the level of fully specific physical arrangements that causation disappears. Take a volume of gas in a container with a safety valve that opens at temperature T (consistent with many specific arrangements of particle movements). T is the cause of the valve opening, not the too specific particle movements. But those particle movements can still cause the precise trajectory of the valve’s opening. Again, the particle movements aren’t maximally specific: the bonding properties of the gas’s molecules, the molecular structure of the valve’s inner surface, and so on, will still be variably realized at the level of fully specific physical arrangements.

C The Argument for Physicalism Re-Examined

18. But now we seem to have cut off the branch we’re were sitting on. We can’t any more assume that ‘Premise 2. All physical effects have full physical causes.’ I have now argued that causes are to be found at a higher level than basic dynamics, and so that there is no immediate sense in which unreduced mental causes are ‘of a kind (also) found outside sentient/living bodies’. Of course, if we are physicalists, we will assume that mental causes are not metaphysically independent of their physical grounds. But I was after a prior argument for physicalism, one that hinged on the thought that if you look back in time from eg a behavioural effect P, then you will always be able to locate full causes at the level of physical realizers, as well as at the level of mental antecedents. And now I am denying just that.

19. One possible response would be to query whether the thermodynamic view of causation really does undermine the claim that we will always be able to locate full causes at the level of physical realizers. For it’s not clear the position I’ve defended rules out mental and other macroscopic cause being type identical with physical facts, and so on that account themselves physical. Advocates of macroscopic mental causation (like Yablo and Menzies) are often taken to be saving non-reductive physicalism, by showing how variably realized mental and other macroscopic facts can be full-fledged causes in their own right. But surely variable realizability requires not that

(I) there are some physical differences between the Ps that determine M, but

(II) there is no distinctive common physical feature of the Ps.

After all, temperature is macroscopic w.r.t. specific particle arrangements that realize it, yet is surely the paradigm of a physically reduced quantity. This shows that macroscopic causes are one thing, and non-reducibility another.

20. I’m inclined to argue that not only can macroscopic causes be physically reducible, but that they must be. Suppose there were no common characteristic, specifiable at the physical level, to all the instances of a given mental M (cf mean kinetic energy as a common characteristic of given temperature T). Then it would seem that there would be no possibility of a uniform explanation of why all the different micro-realizers P of M tend to produce P\*, say. So at the physical level there would seem to be a massive coincidence, with all the physically heterogenous Ps each conspiring to yield P\* in their different ways.

21. All true enough, but in the present context this puts the argumentative cart before the horse. I’m here assuming that M isn’t some metaphysically independent dualist cause; it’s still physically realized, and so there has to be a story about how those realizers all generate the same result. But why are we ruling out that M is a metaphysically independent dualist cause? We had an argument that hinged on ‘All physical effects have full physical causes.’—but we can’t defend that premise by assuming a conclusion that can only be reached once it is given, namely, that M has micro-realizers that need to be co-ordinated. Why suppose we have micro-realizers at all? What’s ruling out full-on interactionist dualism?

22. I’m not sure that we are cheating here. Perhaps we can see the requirement that M have co-ordinated micro-realizers as following from Premise I (it has physical effects) plus what we have discovered a posteriori about the nature of causation. So the idea would be interactionist dualism may be conceivable, but that it’s not metaphysically possible because of the nature of causation. (Still, perhaps this hasn’t ruled out an interactionism based on some metaphysically special mental schmausation . . .)

23. Let’s go back to the idea that there are no special force fields in basic physics. I take the case for this to be untouched by all the above. (Alyssa Ney, when she talked to you on this topic last year, expressed doubts about this, but I am not sure I share them.) Perhaps then we should forget about causes and instead rephrase the earlier argument in terms of ‘determination by law’, with the crucial completeness premise as:

Premise 2 All physical occurrences are determined by prior physical initial conditions in accord with physical law (to the extent they are determined).

But how would the rest of the argument go?

Premise 1 Mental facts determine physical occurrences by law.

Premise 2 All physical occurrences are determined by prior physical initial conditions in accord with physical law.

Premise 3 It would be absurd to hold that all physical consequences of mental antecedents are always also determined by law by metaphysically independent physical antecedents.

Conclusion Mental facts must (supervene on) part of the physical antecedents of their physical consequences.

I have my doubts about Premise 3, and even more so about Premise 1.

24. Here is a different idea. Let’s still stick with the idea that there are no special force fields in basic physics and that ‘all physical occurrences are determined by prior physical initial conditions in accord with physical law.’ But rather than trying to rejig the causal argument, can’t we simply observe that this leaves no room for any metaphysically independent mental causes to make a causal difference to physical effects. Once we have a world whose basic dynamics does not involve any special mental fields, then any variation in metaphysically independent mental events will leave the unfolding of the physical realm unchanged.

25. It mightn’t be clear how this argument works. Even if our basic dynamics doesn’t involve any special nervous forces, why shouldn’t the occurrence of some later P\* be counterfactually dependent on some earlier metaphysically independent M? But let’s think about the metaphysics, not the counterfactuals. Go back to the original argument against ‘overdeterminationism’ based on Premise 3. The idea was that if we had metaphysically independent M and P both causing P\*, this would require an unacceptably unsimple theory. Why exactly? Well, if we focus on those cases where M causes P\*, then some determining P will also have to be present in all of them, to respect Premise 2—and this would require some further explanation, such as that M is nomologically co-ordinated with such Ps synchronically, or M and such Ps are joint nomological corollaries of some yet prior common Q, or whatever. In short, if M is metaphysically independent of the physical world, then we will need some Rube Goldberg set-up to keep them in step, since the metaphysics doesn’t preclude an M without a P\*-determining P.

26. This reasoning goes through just the same even if the P required by Premise 2 doesn’t cause P\*, but just fixes it by law. If we suppose that M is metaphysically independent of any P, then we are faced with the need to explain why the prior physical conditions that obtain when it causes P\* are always such as to determine P\*, and it’s hard to see how any explanation of this could be other than absurd. Note by contrast that if M is not metaphysically independent of the underlying physical facts, then these worries simply do not arise. It will automatically be realized by a P that fixes the later occurrence of P\*. And that’s why we need to accept that mental causes are physical.

27. So, to sum up, the idea that the basic dynamics does not involve any special forces is by itself enough to argue that any mental and other special causes must be physical. We don’t need to put it in terms of a need to avoid a proliferation of overdetermining causes. We can simply say that, if mental causes weren’t physical, we would be at a loss to account for the compliance of the basic dynamics with the results of mental causation.