# Causation is Macroscopic but not Irreducible

David Papineau

**[from S Gibb and E Lowe eds *The Ontology of Mental Causation* 2013]**

## 1 Introduction

In this paper I argue that causation is an essentially macroscopic phenomenon, and that mental causes are therefore capable of outcompeting their more specific physical realizers as causes of physical effects. But I also argue that any causes must be type-identical with physical properties, on pain of positing inexplicable physical conspiracies. I therefore allow macroscopic mental causation, but only when it is physically reducible.

## 2 Causation is Macroscopic and not Physically Fundamental

It is widely supposed, at least among philosophers, that causation is grounded in basic dynamical processes. At bottom, according to this supposition, causal relations consist in the way that basic dynamical laws govern the temporal evolution of precise arrangements of fundamental physical particles and fields.

This supposition is often on display in contemporary debates about mental and other 'higher-level' causation. Suppose you think that some mental state M—your wanting to hail a taxi, say—supervenes on physical state P without being reducible to it. (P is more specific than M: it metaphysically necessitates M, but M can be realized by physical states other than P.) Now consider some further physical effect P*—your arm moving—that apparently results from M. Can M really be the cause of P*? Some philosophers deny that it can, on the grounds that M will always be 'outcompeted' as a cause by P. Others assert that M can be such a cause, arguing that there is nothing wrong with both M and P 'overdetermining' the effect P* in such cases. But scarcely anyone queries whether the realizing P will itself qualify as a cause of P*.[1]

Thus Jaegwon Kim, in discussing just this kind of case, insists that

'The question is not whether P should be considered a cause of P*; on anyone's account, it should be' (Kim 1993 p 207, my italics).

(Kim then continues with his familiar query: 'What causal work is left over for M, or any other mental property, to do?')

However, I myself am very doubtful about the claim that Kim takes to be agreed on all sides.

Why does Kim take it to be obvious that P is a cause of P*? Presumably he is reasoning from the above supposition that causation is constituted by the way maximally precise physical arrangements evolve in accord with basic dynamical laws. If P is a full specification of physical initial conditions, and these evolve in line with basic dynamic laws into P*, then of course—of so Kim assumes—P must cause P*.

However there is strong reason to doubt that that causation is constituted by basic dynamical processes. The objection is simple. Causation is asymmetric in time, but basic dynamics is not. So it seems that causation must involve something more than basic dynamics.

---

[1] I take causes and effects to be facts or states of affairs, paradigmatically consisting of some particular possessing some property. Given this, it will sometimes be natural to talk as if properties themselves are causes or effects; but strictly what is meant by such usages are the facts involving those properties.

The basic laws of dynamics determine no direction in time. Take a specification of what happens at each point of spacetime in some closed physical system. Then you can view the 'initial' conditions as evolving into the 'final' conditions in accord with the basic dynamical laws. But you could equally well think of time as 'flowing' in the opposite direction, with the 'final' conditions evolving into the 'initial' ones, again in accord with the basic dynamical laws. In this sense, the basic laws of dynamics don't care which direction is 'earlier' and which 'later'. Accordingly, if you are given a basic dynamical description of a physical system, but not told which temporal direction is which, you won't be able to read this off from the description.

But causation is different. 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.

Of course, this would be trivial if the difference between 'causes' and 'effects' simply consisted in the former occurring earlier than the latter. Thus suppose that the relationship between 'causes' and 'effects' were just like that between the sets of conditions at the two temporal ends of a basic dynamic process, save that the direction of time was independently given, and it was specified that the 'cause' was the earlier set of conditions, and the 'effect' the later set. Then it would scarcely be surprising that we could read off temporal order from information about 'causes' and 'effects'—for such talk would simply add this temporal information explicitly to the temporally neutral dynamic facts.

But it is arguable that 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. It is noteworthy that the techniques exploited by the Bayesian net tradition do not need to assume temporal order in order to distinguish causes from effects. Sufficiently rich correlational information on its own will always determine a causal order among related variables. (This is not the place to go into details. But, to get a flavour, note that the correlation between the joint effects of a common cause will disappear when we 'control' for the common cause—that is, consider separately cases where the cause is absent and where it is present. By contrast, any correlations between the joint causes of a common effect will not disappear when we 'control' for that effect. For more on the asymmetric probabilistic dimension of causation see Glymour, Scheines and Spirtes 1993; Hausman 1998; Papineau 2001.)

There is room here to debate the precise metaphysical relationship between the underlying causal structure and the temporally asymmetric correlational structure which manifests it. But the very possibility of epistemologically distinguishing causes from effects without assuming temporal order suggests that there must be something in the nature of causation that orientates it in time. If so, this means that there must be more to causation than the temporally symmetric structures of basic dynamics.

## 3 Thermodynamics and Causation

It is illuminating to compare causation with thermodynamics in the above respects. As is well known, the second law of thermodynamics is also inexplicable in terms of basic dynamics alone, precisely because it refers to a specific direction in time: later entropy is always greater than earlier entropy within a closed physical system. So an explanation of the second law needs to invoke assumptions that go beyond basic dynamics. In particular, such an explanation needs to posit, in addition to 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 2000).

I take causal asymmetry to have an analogous basis. There is no established way of relating causation to thermodynamics. But the asymmetric correlational structures displayed by causal relationships suggest that causal processes are akin to thermodynamic processes. In particular, it suggests that causation is also is an essentially macroscopic phenomenon, constituted by the nature of past facts together with probability distributions over the maximally specific microstates that can realize given macrostates.

Some readers might be puzzled by the suggestion that thermodynamic processes, and causal relationships along with them, are essentially macroscopic phenomena. I alluded above to the way that thermodynamic processes like entropy increase can be explained in terms of particle physics, together with past facts and probability distributions over microstates. But if such explanations are possible, then don't they show that the macroscopic thermodynamic phenomena can all be reduced to microscopic processes, and so aren't really macroscopic after all?

But it does not work like that. The explanation of thermodynamic phenomena by particle physics does not eliminate macroscopic features, but makes essential use of them. Take a volume of gas that is hot in one half and cold in another. Thermodynamics 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 of all the particles involved, and this would no doubt have told you that the later temperature would be uniform. But this microanalysis would owe nothing to the general principle that almost <u>any</u> system in that initial <u>macrostate</u> would end up at a uniform temperature. (After all, you could have applied an entirely analogous microanalysis to predict the evolution of one of the very unlikely 'rogue' microstates that would <u>not</u> end up with 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 <u>macrostate</u> which is overwhelmingly likely to (be realized by a microstate that will) end up with a uniform temperature. This is why thermodynamics is essentially macroscopic. Without probabilistic information about the way in which given <u>macrostates</u> get realized by microstates, you cannot infer any thermodynamic patterns from microphysics.

Similarly, I suggest, causation is an essentially macroscopic phenomenon. 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.

## <u>4 Intuitions are Irrelevant</u>

Of course, this is not how we think about causation intuitively. The intuitive paradigm of a causal interaction is of one physical object bumping into another and the latter's motion changing. We humans are naturally prone to judge without further ado that in such cases the impact of the former caused the new motion of the latter. (See Michotte 1945/1963.) In line with this, our intuitive conception of causation contains no mention of probabilistic distributions over the microstates that realize different macrostates, and correspondingly our concept of causation sees no contradiction in the idea of causation existing even when such probabilistic distributions do not.

In this connection, consider Ernest Sosa and Michael Tooley's objection to the way David Lewis analyses causation in terms of the 'asymmetry of overdetermination'. Lewis's analysis

is in the spirit of theories that account for causal asymmetry in terms of probabilistic asymmetry. True, Lewis does not put this analysis explicitly in probabilistic terms, but the appeal to probabilistic facts is not far beneath the surface. (Thus note how Lewis appeals to the way that causes typically issue in many independent chains of effects, yet typically only stem from one chain of causes.)

Sosa and Tooley object to Lewis's theory on the grounds that

". . . 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 1993, p 27).

Well, no doubt there are conceivable scenarios which contain causal relations but lack the asymmetric probabilistic structure to which Lewis appeals in explaining causation. And if Lewis's theory were put forward as a piece of conceptual analysis, then such scenarios would suffice to refute it.

But there is no reason to read theories like Lewis's in this conceptual way. Rather, they are better understood as synthetic metaphysical theories, which aim to uncover the nature of causation, not via analysis of our concepts, but through a posteriori investigation of the world we live in. 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 $H_20$. Of course, if it could be established that worlds with causation but no probabilistic structure were metaphysically possible, then this would indeed defeat probabilistic accounts of the nature of causation. But the mere conceivability of such worlds does not show that they are metaphysically possible. 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.

## 5 Autonomous Mental Causes

So far I have argued that it is a mistake to think of causal relationships as being determined by some maximally specific level of physical facts. Rather causation depends on general patterns essentially involving macroscopic properties, where these macroscopic properties will be realized by different arrangements of fundamental physical facts on different occasions.

Over the last couple of decades a number of philosophers have argued that mental facts M are no less causes of subsequent physical results P* than their physical realizers P. On this view, the result P* can be attributed to both of the 'parallel causes' M and P. Of course, nobody wants to view all mental causation as overdetermination by two ontologically distinct causes, like the death of the man who is shot and struck by lightning at the same time. But defenders of the 'parallel causes view' can observe that M and P are not so ontologically distinct, in that M metaphysically supervenes on P. True, M is not identical to P, and so in a sense a kind of 'benign overdetermination' is being posited. But precisely because M supervenes on P, and is not ontologically independent, it is not obvious that there is anything wrong with such benign overdetermination. (See Shoemaker 2001, Pernboom 2002, Bennet 2003.)

This 'parallel causes view' has the virtue of recognizing macroscopic mental facts as causes in their own right. But the points made in the last section open the way to a more radical position. Why shouldn't the mental state M be the cause of P* rather than the physical state

P? If causation derives from patterns essentially involving macrostates, then perhaps it is the mental M that figures in these patterns, not the physical P.

Just this possibility has been explored by some recent writers. (See Menzies 2008, List and Menzies 2009, 2010. Also relevant are LePore and Loewer 1987, Yablo 1992.) Their standard form of argument appeals to plausible counterfactual requirements on causation. Let us suppose that, if C causes E, then

(1) E wouldn't have occurred if C hadn't occurred, and

(2) E would still have occurred if C had occurred differently.

Take the case where you are waving for a taxi. Let the physical effect P* be the movement of your arm. M is your wanting to hail a taxi. P is the definite neuronal arrangement which realizes this mental state. Now, both M and P satisfy clause (2)—your arm would still have moved as long as either M or P occurred, even if they had occurred in a different way. But only M satisfies clause (1)—your arm wouldn't have moved if you hadn't wanted a taxi—where P does not—the absence of just that precise neuronal arrangement wouldn't have stopped your arm moving, for you would still have wanted to wave even if your desire had been realized by a slightly different neuronal state.[2]

So here M is the cause and not P. P is too specific. This analysis is in line with Stephen Yablo's thesis (1992) that causes should be proportional to their effects. Causes must be specific enough for their effects, but no more specific than this requires. In the above example, the neuronal arrangement P is too specific for the effect, but the wanting M is just right.

Note how the example comes 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 both M and P satisfy clause (1)—if you hadn't wanted to, you wouldn't have waved at all, and so a fortiori not just as you did; and if you hadn't had just precise neuronal set-up, you also wouldn't have waved you just as you did. But, with this precise effect P*, only P satisfies clause (2)—while you would still have moved just like that if P had occurred a bit differently, you wouldn't have moved just like that if you'd still wanted to move your arm but this desire had been realized with some different neuronal arrangement.

So now P is the cause and not M. The wanting M is not specific enough to account for your moving just like that, but the neuronal arrangement P is just right.

Does this last example not run against the points made in my first section? There I said that causation is an essentially macroscopic phenomenon, and disappears at the level of maximally specific physical processes. Now I am saying that your neuronal arrangement P can be the cause of your particular movements, rather than your mental state M. However, these claims are not inconsistent. The reason is that the neuronal arrangement P needn't constitute a maximally specific physical state. Just as a given mental state M can be realized by different neuronal arrangements, so too can a given neuronal arrangement be different realized at the maximally specific level of precise fundamental particles and fields. So in both cases there is 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.

[2] Laurie Paul has queried whether M would have been different realized, rather than simply being absent, if P had not occurred. This does sound natural to my ear, but in any case let me simply stipulate that the suggested counterfactual requirement on causation be read this way.

Let us return to the analogy between causation and thermodynamics. Suppose we have a volume of gas in a container with a safety valve. If the gas is heated, there is a temperature T at which the valve will open. This temperature can be realized by the many different sets of particle movements which would yield the requisite mean kinetic energy. Which causes the valve's opening on some given occasion, the temperature T or the specific particle movements which there realize that temperature? Intuition might suggest that it is the particle movements. But if the effect at issue is the opening as such, the approach I am defending argues that it is the temperature and not the particle movements that cause this effect. The particle movements are too specific. We would still have had the opening even if the temperature T had been realized by different particle movements.

Does this mean that the specific particle movements cannot cause anything? That would be an undesirable conclusion. There are more fine-grained effects, such as the precise trajectory of the valve's opening, which will surely be the results of the specific particle movements. Perhaps the valve's opening in that specific manner is due to precise sequence of high-energy particles that impact its inner surface. With the more fine-grained effect, it is the more specific particle movements and not the generic temperature T which is proportional to the cause: if the temperature T had been differently realized, then the valve would not have opened in just that manner.

Note how the causal efficacy of the particle movements does not undermine my claim that causation is essentially a macroscopic phenomenon. Even after we have focused on the definite particle movements, there will be yet further features of the set-up—such as the bonding properties of the gas's molecules, the molecular structure of the valve's inner surface, and so on—that will still be variably realized at the level of fully specific physical arrangements. And it is still probabilistic facts about the distribution of such further realizers that underpin the asymmetric causal relationship between the particle movements and the manner of opening. If we descend to a level where all physical facts are fully determinate, then I say that we lose sights of any asymmetric causal relationships. But this leaves plenty of room for relatively definite physical facts like given particle movements to function as genuine causes of relatively fine-grained effects. We can descend to particle movements without descending to the level of fully specific physical states.

I alluded earlier to the idea that a mental cause M and a physical realizer P might both be the cause of some physical effect P*, via a sort of benign overdetermination. The points made so far in this section argue that this is not a possibility—not on the grounds that such overdetermination per se would generate any unacceptable consequences, but simply because the requirements of proportionality rule out two such causes. Once we have fixed on a specific effect P*, then it can't be that some M and some more specific realizer P are both causes of P*. If M is proportionally 'just right', then the more specific P will violate requirement (1), in that we would still have had P* without P. And if P is proportionally 'just right', then the less specific M will violate requirement (2), in that we wouldn't still have had P* if M had been realized differently. It can't be the case that a more and less specific state are both causally proportional to a given effect. (Cf Loewer and Lepore 1987, List and Menzies 2010.)

Suppose that some mental state M outcompetes its more specific physical realizer P as the cause of some physical effect P*. What then is the relation of the realizer P to the effect P*? Many of those who defend the causal status of M nevertheless retain the idea that the physical realizer P is 'causally sufficient' for the physical result P*. Their thought is that P still causally determines P* even though it is too specific to count as 'the cause' of that result. But from the perspective being defended here, even this seems to concede too much to the intuition that causation is grounded in basic physical processes. Of course we might wish to allow that such specific physical antecedents are nomologically sufficient for the subsequent physical results. But there is no reason to think of this sufficiency as a causal matter, in cases

where the precise physical detail omits any mention of the macroscopic pattern that constitutes the causal relationship.

## 6 Proportionality and Reduction

Does the fact mental states can eclipse their physical realizers as causes of certain effects vindicate the possibility of non-reductive physicalism in the philosophy of mind? This conclusion is typically drawn by those philosophers who stress that proportionality requirements can favour mental states as causes over their physical realizers. But it is by no means clear that it follows.

Reductive physicalism requires the type identity of mental properties with physical properties. Non-reductive physicalists maintain that no such type identities are available. It is important to realize that, in order to establish non-reductive physicalism, it is not enough to show that there are some physical differences present on the different occasions where M is realized. Rather we need to show that there is no distinctive physical commonality present on all those occasions.

When Putnam and Fodor introduced the idea that mental and other special science properties might fail to reduce to physical properties, they weren't just making the weak claim that different instances of these properties will display some physical differences. Rather their idea was that there would be no common physical feature of different instances. There would be nothing physically in common between the different computers that can run a given programme, or between the different organisms across the universe that can think a given thought.

So far in this paper we have been dealing with cases where some mental M is realized by a different more specific physical Ps on different occasions. That is, we have been dealing with cases where distinguishable physical Ps can metaphysically determine the same mental M. This by itself fails to establish the anti-reductive thesis that there is no further physical feature Q which is type identical to M.

To see that this stronger anti-reductive thesis does not follow from the fact that M is determined on different occasions by distinguishable Ps, we need only consider a thermodynamic example once more. Take the case where the gas reaching temperature T opens the valve. Temperature T can be realized by many distinct arrangements of specific particle movements. But it does not follow that there is no further physical property which characterizes all the instances of T. And of course in this case there is. All the instances of T involve arrangement of particles with the same mean kinetic energy. And it is precisely this common physical feature which allows the possibility of a uniform thermodynamic explanation of why the valve will open at that temperature. The probability distribution over the possible microstates that realize that mean kinetic energy implies that the valve is overwhelmingly likely to open at that temperature.

This is surely the paradigm of a type-type reduction. We identity temperature with some common physical feature specifiable in terms of particle movements, namely a given mean kinetic energy, and thereby explain patterns involving temperature in terms of particle physics.

But as well as being the paradigm of a type-type reduction, this is also a case where proportionality considerations point to the macroscopic temperature as the cause of the valve opening, rather than the more specific particle movements which realize it on given occasions. I infer that there is nothing in the idea of macroscopic facts being proportionate causes to rule out fully reductive physicalism.

Recall Putnam's famous example of the square peg and the round hole (1975). Putnam argued that the properties of squareness and roundness will be a much better explanation for why the peg does not fit in the hole than any detailed specification of the quantum mechanical arrangement and properties of relevant bodies' molecules. Quite so. It is the squareness and roundness that are proportional to the peg's failure to fit, not the very specific molecular arrangements that realize these properties.

But this does not mean that squareness and roundness are not physically reducible. We can still specify features of their molecular arrangements that will be common to all square pegs and round holes, and we can appeal to the so-specified features to explain at that level why square pegs don't go into round holes. Here again we see that it does not follow from the causal dominance of macroscopic facts over their more precise realizers that those macroscopic facts must be physically irreducible.

These examples manifest a typical set-up in physics. Some macroscopic property common to many microscopically distinguishable states can be identified with some common feature of those microscopic realizers, and this common feature then accounts for the way that the macroscopic property features essentially in some general pattern.

## 7 Against Unreduced Causes

The last section showed that macroscopic causes are one thing, non-reducibility another. There are plenty of cases where macroscopic properties can feature as proportionate causes of certain physical effects, and thereby causally eclipse their more specific microphysical realizations, and yet these macroscopic properties are fully reducible to some common physical feature of their microscopic realizations.

I now want to argue that macroscopic causation is not just consistent with physical reducibility, but that it positively requires this.

I have argued in this paper that asymmetric causal relations derive from probabilistic facts about the way in which macrostates are realized at the micro-level. This picture assumes that each macro-cause corresponds to some constraint specifiable at the micro-level, in the way that temperature corresponds to mean kinetic energy. The probabilistic facts about the different ways this condition can be satisfied by precise microstates then accounts for the asymmetric causal patterns involving macrostates.

If this is the right general story about causation, then it is hard to see how macroscopic causes can fail to be physically reducible. Their very nature as causes will derive from their type-identity to some physically specifiable constraint, for it will only be in virtue of this identity that they systematically generate their effects.

What options are open to non-reductive physicalists here? There seem to be two ways they might go. First, they might argue that, when a given macroscopic cause is variably realized, it generates its effects via different causal processes at the physical level. Alternatively, they might argue that there is no need to invoke any casual processes at the physical level to explain how a variably realized macro-cause generates its effect. However, neither of these options seems at all attractive.

To bring out the difficulties here, note that proportionate causation involves an element of generality. Recall that our two requirements for C to cause E were that

(1) E wouldn't occurred if C hadn't occurred, and

(2) E would still have occurred if C had occurred differently.

Clause (2) here tells us that in other similar circumstances where C occurs, E will occur too. On other similar occasions where I want to hail a taxi, my arm still moves. And clause (1) tells us that when C doesn't occur in similar circumstances, E will fail to occur too. On other similar occasions where I don't want to hail a taxi, my arm doesn't move. In short, there is a general co-variation of C and E in similar circumstances.

Now, the problem facing non-reductive physicalists is to explain why C and E should so co-vary if there is no uniform physical condition corresponding to C which can account for this. The answer is obvious if C can be identified with some physical condition which systematically generates the result E. But in the absence of any such identification, non-reductive physicalists seem to face a challenge.

The first non-reductive response to this challenge would be to hold that the different realizations of C give rise to E via <u>different</u> causal processes. This is probably how non-reductive physicalism is normally understood. On different occasions when people want to hail a taxi, their desire is realized by different physiological arrangements—but each of these different physiological arrangements has the causal power to produce an arm movement.

The trouble facing this option is that we have been given no account of why the different causal processes that realize C should all alike be ones that give rise to E. For all that has been said so far, this looks like a mystery. If the processes at the physiological level are all so different on different occasions of desiring to hail a taxi, why ever should they all be followed by E?

To focus this issue, it will be helpful to consider an inorganic example, as there are features of mental phenomena that can obscure the difficulty at hand, in various ways to be considered below. Let us imagine that the water from a certain lake seems to have a distinctive power to destroy rubber. But when we look into the mechanism, we find no common causal process. In one case, the water contains rubber-eating bacteria. In another, the water turns out to be highly acidic. In yet another, there are high levels of ozone in the water and this produces a rubber-destructive agent. And so on. In each of the cases that we examine, we find a physical explanation for the rubber's deterioration, but the explanation is different in each case.

I take it that this story does not hang together. If we really came across a case like this, and discovered a different mechanism in each case, we would surely conclude that it wasn't a genuine causal relationship after all, and that the observed pattern was just a coincidental feature of the cases so far observed. I think that we should have the same reaction to the suggestion that some mental C can produce a physical effect E via different causal mechanisms on different occasions. In the absence of any further information, it seems incredible that nature should work like this.

Let me now consider the alternative non-reductive response to the challenge of explaining why some C and E should co-vary if there is no uniform physical condition corresponding to C which can account for this. The alternative non-reductive option would be to deny that we need <u>any</u> causal accounts at the physical level for macroscopic causal processes. Now the idea is not that <u>different</u> causal processes account for the C-E link in different cases, but that there are <u>no</u> further causal stories to be told at the physical level at all.

But this too looks like mystery-mongering. Remember that we are exploring the possibility of non-reductive <u>physicalism</u>. It is not as if we are positing some ontologically independent realm of mental causes with brute powers to produce physical effects. Rather, we are taking it that mental and other macroscopic causes metaphysically supervene on the physical facts—nothing more is required for their presence than those physical facts. But then causation

without a physical explanation looks like a conspiracy. On different occasions C is realized by different physical microstates, and somehow these all evolve into later microstates that determine E. But there are no conditions satisfied by the initial microstates that might account for their all evolving into states that determine E. I see no reason to accept that there are macroscopic causal patterns which correspond to no causal patterns at the physical level in this way. Imagine that temperatures supervened on molecular motions, and that certain temperatures produced certain regular effects, but that there was no uniform story available at the molecular level of why this should be so. This doesn't seem the way that things work in our world.

Note that I am not accusing non-reductive physicalism of any outright inconsistency. There is nothing contradictory in the idea that the physical realizers of some C should all just happen to eventuate in some E, either via different causal paths at the physical level, or via no such casual paths at all. But I take it that our experience shows us that the world just doesn't work like that. Macroscopic causal patterns do not depend on massive coincidences at the physical level. Rather any macroscopic cause corresponds to a common physical condition satisfied by its realizers, and there is a physical story to be told about why this condition gives rise to the relevant effect.

## 8 Causes not Laws

In a number of previous papers (Papineau 1985, 1992, 2010) I have offered a similar argument against non-reductive physicalism. However those earlier papers focused on laws, not causes. My earlier arguments owed nothing to the way that asymmetric causation depends on probability distributions over the microstates consistent with a given macro-condition. Rather I simply appealed to our supposed knowledge of the way general laws of any kind depend on uniform physical processes, arguing that it would be incredible that there should be a law involving physically supervenient properties, yet no uniform physical account of the way that the physical realizations of the initial condition evolve into physical realizations of the final condition.

However, this line of argument can be criticized for its appeal to the unexplained notion of a 'uniform' physical account. Take the full range of nomologically possible physical microstates that can realize the relevant initial macrostate. Now suppose that those initial conditions evolve according to the basic dynamical law (Newton's second law in a classical context, or Schrödinger's equation in a quantum context). The upshot will be that this range of initial microstates will be shown to end up in a range of later microstates that determine the relevant final macrocondition. On what grounds do I say that this is not a 'uniform' account of the original macroscopic law? It can't just be that the basic dynamic law is being applied to a range of distinguishable microstates. That would rule out pretty much anything as a uniformly explained law, including any laws of thermodynamics, chemistry or planetary motion.

Yet in my earlier papers I offered no other account of what might render a physical account non-uniform. Given this, it is unclear what force there is to my insistence that it is 'incredible' that there should be macroscopic laws that lack a uniform physical account. In the absence of some further explanation of what counts as uniform, what exactly is it that I say I find incredible?[3]

Of course, this is not an objection that can happily be made by those who want to define themselves as non-reductive physicalists, since they too will need to appeal to a notion of 'uniform'. This is because they want to insist that the application of the basic dynamic law to the collection of initial conditions in not a uniform physical reduction, but a derivation that

[3] Barry Loewer has pressed me on this point in conversation on a number of occasions.

covers a physically heterogenous range of cases. Still, the point remains that, without some further account of what counts as a 'uniform' physical process, it is unclear what substance there is to my dispute with non-reductive physicalism. It seems as if there may be nothing at issue when I claim that macroscopic laws must be physically reducible, and they deny this.

I take this paper to add substance to my position by focussing on causal processes. I no longer wish to argue that all laws are physically reducible—I concede that there may be no good sense in which all macroscopic laws must have a uniform physical reduction. Rather my focus is now specifically on asymmetric causal patterns, and my claim is that for any such causal pattern, there will be a constraint specifiable at the physical level common to all realizations of the cause, and that a probability distribution over the microstates satisfying this constraint will play a part in explaining why the effect follows. This is what I mean by a uniform physical explanation for a causal pattern. So my present thesis is that there are no causal patterns in our world that lack uniform explanations of this kind.

## 9 Does Functionalism Help?

It might have occurred to some readers to wonder whether a functionalist account of mental states might not help to explain how a given mental cause may produce its results via different physical-level causal processes on different occasions. If mental states are defined functionally as states which produce certain effects, then won't such variable causal mechanisms be just what we would expect? Suppose the mental state of wanting to hail a taxi is defined as a state that will produce arm movements or similar signals. There may well be lots of different physiological states that satisfy this requirement. But it will scarcely be 'incredible' that they should all alike give rise to arm movements—for it is just this tendency that qualifies them as realizations of wanting to hail a taxi in the first place.

However, the appeal to functionalism does not help. As is well known, there are two very different versions of the functionalist thought that mental states can be 'defined' as states which produce certain effects. Once they are clearly distinguished, we can see that neither of them helps non-reductive physicalism to explain how one state can cause another via different causal processes. The impression that functionalism helps with this problem only arises if the two versions are run together.

The first version—realizer functionalism—is a thesis about how the reference of mental terms is fixed. On this view, mental states are physical states that are identified via their connection to certain causes and effects. For example, the term 'desire to hail a taxi', applied to some person, is to be understood as referring to that physical state which causes appropriate arm movements in that person. This term might thus refer to different physical states in different people, just as the term 'your watch' might refer to different devices when different people are being addressed.

This view does nothing to explain variably realized causes, for the simple reason that it does not trade in variably realized states of any kinds. The only states it countenances are ordinary physical states—such as the physical state which causes my arm to move, say—and these physical states cause their effects in an ordinary uniform manner. True, different physical states may well be picked out by the same mental word, in virtue of producing some common effect—but there is nothing here to suggest that any given such state produces its effect via different routes on different occasions.

Then there is role functionalism. This does recognize states which are variably realized at the physical level. Role functionalism takes mental terms to refer, not to the first-order physical states that have certain specified causes and effects, but to the second-order states of having some first-order state that plays that causal role. On this view, mental terms will have the

same referent even when applied to differently constituted beings: they refer to the second-order state shared by all beings who instantiate the relevant causal role.

We can usefully bring out the difference between realizer and role functionalism by thinking of a term like 'dormitive virtue' as applied to sleeping pills. The realizer option would take this term to refer to the narcotic chemical constituent present in whichever sleeping pill is under discussion. The role option, by contrast, would take the term to refer to the property common to all sleeping pills, namely, their tendency to produce sleep by whatever means.

Now, role functionalism does arguably give us variably realized states. Just as a tendency to produce sleep can be realized by different chemical processes, so can a tendency to move ones arm be realized by different physiological processes. The trouble is that, if mental states are like tendencies to produce sleep, then surely they are disqualified as causes of the effects that constitute them. A tendency to produce sleep isn't sufficiently distinct from the sleep itself to qualify as its cause. Similarly, if a desire to hail a taxi constitutively requires appropriate arm movements, it isn't distinct enough form the arm movements to cause them.

So whichever way we turn functionalism, it doesn't give us causes which produce their effects via non-uniform physical processes. Realizer functionalism gives us causes all right, but they operate in a physically uniform manner. Role functionalism gives us variably realized states all right, but they aren't causes of the relevant effects.

## 10 Selectional Properties

Perhaps a different kind of functionalism can account for variably realized causes. Rather than considering states that are defined or constituted by a causal role, let us instead consider states that are functional in the sense that they have been selected to play some causal role.

The puzzle I have been pressing so far is how some physically supervenient putative cause C can regularly co-vary with some putative effect E if there is no common feature at the level of its physical realizations to account for this. In the absence of any such commonality at the physical level, it seems mysterious that E should generally follow.

But now suppose that the instances of C have been selected because they produce result E. That is, they occur as the result of some selection process that favours items that produce E. Then the puzzle would be explained. There would be an explanation for why C generally leads to E even though there is no uniform explanation at the physical level. E generally follows because different instances of C have been selected to produce precisely that result.

To illustrate, consider the simple example of thermostats in electrically controlled domestic hot water heating systems. Any such system contains a thermostat which stops the heating once the water reaches some set temperature. But these thermostats involve a range of different mechanisms at the physical level, including bi-metallic strips, expansion gases, mercury bulbs, and thermocouples. Yet there is clearly no puzzle here as to why these different kinds of thermostat always produce the same effect of stopping the heating. Their mechanisms have been selected by the heating designers precisely in order to produce this effect.

So maybe this is a model for unreduced causes. Take the property, in a heating system, of containing a thermostat. Let us suppose that this property does not constitutively involve the effect of stopping the heating, and so is a candidate for causing that effect. Won't this now amount to a case where this physical effect is caused by a variably realized property, namely

the property of containing a thermostat?[4]

Certainly the counterfactuals seem to vindicate this claim. It is the generic presence of a thermostat per se, rather than the specific mechanism that realizes it in a given case, that comes out as proportional to the effect of stopping the heating. If (1) there hadn't been a thermostat, the heating wouldn't have stopped. And if (2) the thermostat had been realized differently, the heating would still have stopped. By contrast, the specific mechanism does not seem proportional. While it is true (2) that a differently realized bimetallic strip would still have stopped the heating, it isn't true (1) that if there hadn't been a bimetallic strip, the heating wouldn't have stopped—because in that case a different design of thermostat would no doubt have done the job instead.

Of course mental systems are not designed by intelligent engineers in the way that heating systems are. But, to the extent that they are designed by phylogenetic and ontogenetic selection processes, the same moral will apply. These selection processes will ensure that there is some mental component available to produce a given effect, but the precise mechanism that does this may vary from case to case.

Thus consider pain across different species. Intergenerational genetic selection will have ensured that all organisms have some mechanism that responds to bodily damage by seeking to avoid the source of the damage. But it may well have lit on different things to do this job in different species.

Nor is the point restricted to the way that the products of intergenerational genetic selection can vary across species. Humans and other complex animals are sophisticated learning machines that embody a hierarchy of processes that operate to preserve items that produce such-and-such effects. The items selected may well be physically different in different individuals, or even in the same individual at different times, but this won't matter to the selection mechanisms, provided they produce the reinforcing effects. The state which leads me to hail a taxi when one is needed may be quite differently realized in me and in you, but we are both likely to possess some such state.

Just as with the thermostats, proportionality considerations again suggest that such selectional mental states can qualify as variably realized causes of physical effects in their own right. Consider again the state of wanting to hail a taxi, and the effect of my arm moving. If (1) I hadn't wanted to hail a taxi, my arm wouldn't have moved. And (2) if this desire had been realized differently, my arm would still have moved. But now consider the specific brain state that realizes the desire in me. While it is true (2) that if this brain state had been realized differently my arm would still have moved, it isn't true (1) that if I hadn't had that brain state, my arm wouldn't have moved—because in that case a different brain state would have been selected to move my arm instead.

## 11 Too Many Causes

This might all now look like good news for unreduced mental causes. However, I think that

[4] The most natural way of construing selectional properties like being a thermostat is as constitutively involving some past history of selection. So understood, selectional properties arguably won't constitutively involve their effects as role properties do—something can be selected to do F and yet have no tendency to do F in the future. Even so, selectional properties might be held to be ineligible as causes on the different grounds that historical provenance cannot matter to causal significance. I shall not press this particular worry, however. Perhaps it can be avoided by construing selectional properties in some way that disconnects them from their history. But even if it can be so avoided, the causal status of selectional properties would still be open to the more the more fundamental objection made below.

appearances are deceptive. Despite the points made above, there is a strong reason to doubt that selected mental items, and indeed selected items generally, can feature as non-reduced causes in their own right. To the extent that the proportionately counterfactuals argue differently, I say that these counterfactuals are misleading as to causal structure.

Let us ask why certain physical states are selected to play a certain role in a cognitive structure or other designed system. The answer is that these states are apt to cause some specific effect, and the relevant selection mechanism favours items with this feature. However, if this is the reason why these physical states are selected, it rules out the more generic variably realized selectional state from also causing that effect.

Take the thermostat example again. If a heating engineer chooses to put a bimetallic strip into the electric circuit, this is because this item will cause the circuit to break when the temperature rises. It is precisely the causal status of this item that renders it suitable for the engineer's purpose. But this then undermines the thought that having a thermostat per se causes the circuit breaking. This generic property is common to different kinds of circuits, and in each of these the breaking is caused by a different mechanism. Having a thermostat itself does not cause the result, for having a thermostat depends on being in some more specific state which does cause the result.

The same point applies to selected mental causes. Why have phylogenetic and ontogenetic selection processes picked certain brain states for the wanting-to-hail-a-taxi-role? Because those brain states get activated when a taxi is needed and they then cause arm waving or similar movements. The relevant selection mechanisms will favour just those brain states that have this causal profile. But this again argues that the selectional property of wanting-to-hail-a-taxi cannot itself cause anything. Wanting-to-hail-a-taxi involves being in a brain state which itself causes arm waving and so on. It was because this brain state already caused this result, so to speak, that it was selected. Given this, it makes little sense to think of the generic selectional state as also causing the result.

Note that this analysis does not appeal to some unthinking intuition that more specific physical states always casually outcompete any more generic states that supervene on them. As the earlier sections of this paper will have made clear, I regard this intuition as fundamentally misguided. Rather I have a more particular objection to viewing generic selectional states as causes. This objection derives from the structure of the selection processes that account for such selectional states, and in particular which explain how they can produce uniform effects despite being variably realized. Selection processes operate on causal facts. Their workings hinge essentially on the causal properties of the items selected. They preserve items that cause certain effects. (See Papineau 2003). This is why we are forced to accept that it is these realizing items that cause those effects, and not the generic selectional states that supervene on them.

What about the counterfactuals? As we saw, they do seem to indicate the generic selectional states as causes, in preference to the more specific mechanisms that realized them. If we hadn't had the generic state, we wouldn't have had the result. By contrast, it's not true that the result wouldn't have occurred if we hadn't had the specific realizing mechanism,—since in that case some other item would no doubt have been selected to produce the result instead. So it looks as if the generic state is proportional to the effect, rather than the specific realizer.

However, I take this to be analogous to the many familiar cases where the counterfactuals fail match causal structure because of back-up arrangements. When I make an assassination plan with a contingency arrangement (for example, a back-up assassin lest the first one fail), it is the whole plan that is proportional to the death of the prisoner, not the shooting by the first assassin. Yet it is that first assassin that caused the death, and the back-up assassin played no causal part. Similarly in the cases at hand: it is the selectional state of being designed for

some end that is proportional to the effect, but the specific mechanism that fulfils the design that actually causes it.

## 12 Explanation is Different

Variably realized selectional states may not cause physical effects, but this does not mean that they cannot be used to explain them. It will be worth clarifying this issue, in order to forestall any inference from the explanatory significance of variably realized states to their causal efficacy.

In this connection, note first that we can often refer to genuine causes indirectly, by citing variably realized selectional states like wanting to hail a taxi. This is possible because we can use descriptions involving the selectional states to construct variable names for the genuinely causal physical states. The way this works has already been discussed under the heading of 'realizer functionalism'. We saw there how we can read 'dormitive virtue' as referring to the specific chemical property present in whichever sleeping pill is under discussion. Similarly, we can understand 'wanting to hail a taxi' as referring to the specific brain state that makes the relevant subject's arm move. So understood, claims like 'he fell asleep because he took a pill with dormitive virtue', or 'his arm waved because he wanted to hail a taxi', will state causal truths.

Not only will such claims state causal truths, but they can also be explanatory. Explanations of particular facts need to name causes. But they can do so indirectly, using descriptions involving the selectional status of those causes. As long as this mode of reference shows us how those causes fit into patterns that can be used to anticipate and control, the attribution of causes will be explanatory.

Thus it can certainly be explanatory to say that someone fell asleep because he took a pill with a dormitive virtue (as opposed to having had a very tiring day, say). Similarly, it can be genuinely explanatory to say that someone's arm waved because they wanted to hail a taxi. (Not all indirect references to causes are explanatory. It is not explanatory to say I fell asleep because I was caused to fall asleep. We need to cite the cause in a way that fits it into a practically significant pattern.)

Selectional states of all kinds are very commonly cited in explanations. I might explain the high temperature in the room by the setting on the thermostat, or the improved performance of my car by its new carburettor. I may have no idea of the actual mechanisms in either case. But knowledge of design properties tells me how the relevant items will work and so suffices for explanatory purposes. Similarly with mental explanations. The states we cite may be variably realized selectional states which are not themselves causes, but they can be genuinely explanatory for all that.

## 13 Causal Closure

The principle of the 'causal closure of the physical' has played a significant role in recent philosophy of mind. According to this principle, every physical effect must have a physical cause. It is this principle that lies behind the widespread modern acceptance of physicalism. (Papineau, 2002.) It allows us to argue that any non-physical realm can only be epiphenomenal, since it would generate an unacceptable overdetermination of physical effects to attribute them to non-physical causes in addition to the physical ones already guaranteed by closure.[5]

[5] The literature displays different uses of this argument. Some use it only to rule out forms of dualism on which the mental realm does not even supervene on the physical. But others, most prominently Jaegwon Kim, also use it to argue against 'non-reductive physicalisms' that respect supervenience but

It is tempting to infer the falsity of the closure principle from the possibility of macroscopic causes. If, as proportionality considerations argue, macroscopic causes can outcompete the more specific realizers as causes of certain physical effects, does this not show that those physical effects at least will have macroscopic causes rather than physical ones, and therefore that the physical realm is not causally closed? (Cf List and Menzies 2010.)

Rejecting causal closure would have a cost. Without a principle of causal closure, we would be left with no argument against interactive Cartesianism and other strong forms of dualism. Fortunately, the possibility of macroscopic causes does not refute causal closure. This would only follow if macroscopic implied non-physical. I have argued that it does not. To repeat my standard example, temperature is a macroscopic property, but it can be type-identified with the physical property of mean kinetic energy.

Moreover I have argued that, not only is macroscopicity consistent with physicality, but that macroscopic causation positively requires physicality. If a macroscopic cause cannot be type-identified with a physical property, we can't give a uniform explanation of why the same physical effect always follows from its different realizations. Nor does it help to appeal to selection processes to explain this, for it is built into the nature of selection that the relevant effects are caused by realising mechanisms, rather than by the generic selectional states that these mechanisms determine.

So my overall analysis reinforces the causal closure of the physical. Certainly many physical effects should be attributed to macroscopic causes rather than their more specific realizers. But these macroscopic causes will still always be physical, thus upholding the principle that every physical effect must have a physical cause, and leaving the argument against dualism intact.

## 14 Mental Causes

One last point. In the latter half of this paper I have been arguing that variably realized mental states cannot be causes. But this does not of course mean that mental states as such can never be causes. For there remains the possibility that some mental states can be type identified with physical states, in the way that temperature is type identified with mean kinetic energy.

I have paid little attention to this possibility so far, given that my main concern has been to establish that variably realized states cannot be causes. But the physical type identity of at least some mental causes is a serious option. Remember that type-identity does not require that there can be no physical differences between the bearers of a given mental state, just that there should be some physical commonality which might explain why the state regularly produces certain effects.

It seems very likely that a wide range of mental states are so uniformly realized within humans, and indeed across many of the other taxa to which we belong. For example, there is every reason to suppose that the pain mechanism is uniformly realized across humans and similar mammals. Again, many sensory mechanisms can be expected to be physically uniform in this way. Perhaps the basic mechanisms of learning and reasoning will also be uniformly realized in all humans, even if not in other species. Provided that we understand our mental terms for these categories as indexed to the appropriate range of species, we can

---

deny type identity. The latter form of argument assumes that any kind of overdetermination is unacceptable, even when one cause supervenes on the other. The former can allow such supervenient overdetermination, and need assume only that overdetermination by metaphysically distinct causes is unacceptable.

read them as referring to physically reducible types, and hence to fully causal states.

On the other hand, I accept that many other mental states will be variably realized across humans. These will be states which derive from ontogenetic selection processes. For example, I would expect wanting to hail a taxi to be variably realized within humans, and even perhaps within individuals. States like these will thus not be causally efficacious, even though they can be explanatory significant in the way explained above.

It is an intriguing question which states are which. For everyday explanatory purposes the difference may not matter much, given that both kinds can equally be invoked in explanation. But the contrast will be significant for cognitive science. Investigation of the physical nature of physically reducible states could bring important scientific benefits, but a similar investigation of variably realized states would inevitably be fruitless. Cognitive science thus needs to know which mental states are causal in their own right, and which play only an explanatory role.[6]

References

Albert D. 2000 Time and Chance Cambridge, Mass: Harvard University Press

Bennet, K. 2003 'Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It' Nous 37

Hausman, D. 1998 Causal Asymmetry Cambridge: Cambridge University Press

Spirtes P., Glymour, C. and Scheines, R. 1993 Causation, Prediction and Search New York: Springer-Verlag

Kim, J. 1993 'The Nonreductivist's Troubles with Mental Causation' in Heil J. and Mele A (eds) Mental Causation Oxford: Oxford University Press

LePore, E. and Loewer, B. 1987 'Mind Matters' Journal of Philosophy 84

List, C. and Menzies, P. 2009 'Non-Reductive Physicalism and the Limits of the Exclusion Principle' Journal of Philosophy 106

List, C. and Menzies, P. 2010 'The Causal Autonomy of the Special Sciences" in Macdonald, C. and Macdonald, G. (eds) Emergence in Mind Oxford: Oxford University Press

Menzies, P. 2008 'Causal Exclusion, the Determination Relation, and Contrastive Causation' in Kallestrup, J. and Hohwy, J. (eds) Being Reduced: New Essays on Reductive Explanation and Special Science Causation Oxford: Oxford University Press

Michotte, A. 1946/1963 The Perception of Causality (English translation 1963 by Miles, E and Miles T.) New York: Basic Books

Papineau, D. 1985 'Social Facts and Psychological Facts' in Currie, G. and Musgrave, A. (eds) Popper and the Human Sciences Dordrecht: Nijhoff

Papineau, D. 1992 'Teleology and Irreducibility' in Charles, D. and Lennon, K. (eds) Reduction, Realism and and Explanation Oxford: Oxford University Press

[6] I would like to thank Eleanor Knox, Barry Loewer, Laurie Paul and David Yates for valuable comments on earlier drafts.

Papineau, D. 2001 ‘Metaphysics over Methodology—or, Why Infidelity Provides no Grounds to Divorce Causes from Probabilities’ in Galavotti, M.-C., Suppes, P. and Costantini, D. (eds) Stochastic Causality Stanford: CSLI Publications

Papineau, D. 2002 Thinking about Consciousness Oxford: Oxford University Press

Papineau, D. 2003 ‘Causation as a Guide to Life’ in my The Roots of Reason Oxford; Oxford University Press

Papineau, D. 2010 ‘Can any Sciences be Special?’’ in Macdonald, C. and Macdonald, G. (eds) Emergence in Mind Oxford: Oxford University Press

Pereboom, D. 2002 ‘Robust Nonreductive Physicalism’ Journal of Philosophy 99

Putnam, H. 1975 ‘Philosophy and Our Mental Life’ in his Mind, Language and Reality Cambridge: Cambridge University Press

Shoemaker, S. 2001 ‘Realization and Mental Causation’, in Gillett, C. and Loewer, B. (eds) Physicalism and its Discontents Cambridge: Cambridge University Press

Sosa, E. and Tooley, M. 1993 ‘Introduction’ to their edited volume Causation Oxford: Oxford University Press

Yablo, S. 1992 ‘Mental Causation’ Philosophical Review 101