

## Should We Give Up on Reductive Physicalism? *Paul Sperring*

Supposing you were a physicalist in the late 1950s, early 1960s, and supposing you were Australian too<sup>1</sup>, it is highly likely you would have thought that mental properties could be reduced to physical properties. Now, suppose you are a contemporary philosopher of mind and suppose further that you are also of a physicalist stripe. Will you be inclined to think that mental properties are reducible to physical properties? It's by no means certain. These days physicalists fall into two, broadly conceived, camps: (i) the reductionist physicalists who think that minds (or mental properties, or states or events<sup>2</sup>) can be reduced to brains (or something smaller) and; (ii) the non-reductive physicalists who think that minds are not straightforwardly reducible to some lower level set of physical properties.

In truth if one were to carefully classify all the physicalist positions in contemporary philosophy of mind we would need distinctions of a much finer grain than this story suggests. For the purposes of this paper, however, those philosophers who have thought that mental properties can be reduced to lower level properties will be lumped together (and called 'reductionists') and those philosophers who, although embracing physicalism, have thought that mental properties in principle defy reduction to something lower down will also be lumped together (and called 'non-reductive physicalists'). As far as name calling goes these days 'reductionist', in some hands, takes a bit of beating as a philosophical insult.<sup>3</sup> The problems with reductionism have been well documented of course, and I will focus on one particular attack. But I also want to say something about what might be worrying about the non-reductive physicalist's position.<sup>4</sup>

### *The Type-Identity Thesis*

One of the most famous (if short-lived) reductionist physicalist views is known as the Type-Identity thesis. Various philosophers in the 1950s argued that mental properties are identical to physical properties – thoughts, sensations, feelings and so on (mental 'types'<sup>5</sup>) *are just* brain processes (physical 'types'). So when persons are said to have some mental property, say the property of being in pain, then analysis will reveal such a property to be a property of their brains, and nothing more.

Perhaps you are struck by the *prima facie* plausibility of such a position. Certainly, these days, when we think of minds we automatically assume a role for the brain, knowing what we know about the close relationship between the two. So when we come to examine how the mind and the brain are related we often think about it in terms of correlation or cause. By correlation we just mean that wherever we have some mental property it is always accompanied by some physical property. Now, mere correlation explains nothing of course. We want to know *why* every mental property is accompanied by a physical property. So perhaps we think about the relationship in terms of causation. All mental properties, we might say, are *caused* by physical properties of the brain. However, if we stop here then we still face a problem. How, we ask, do physical properties give rise to mysterious non-physical properties, and how, if those mental properties are non-physical do they exert any influence over

the properties of the brain (this is the well-known interaction problem that Princess Elisabeth of Bohemia brought to Descartes' attention in the Seventeenth century, although the debate was couched in terms of substances rather than properties)?

One way out of this problem is to say that the reason that mental properties are always accompanied by brain properties is because mental properties *just are* properties of the brain. There is also no residue to explain causally – i.e. how mental properties *arise* as a result of brain goings on – since we only have one type of phenomenon, at bottom. If A and B are identical then A can't be the cause of B, and there is no mystery how B can cause things if there is no mystery about how A causes things (and if A is a brain property then ultimately physics will tell us how that does what it does<sup>6</sup>).

Some of the advantages of accepting the type-identity thesis ought to be pretty clear. Firstly, the problem Descartes wrestled with, how minds and bodies could be in causal communion, seems to melt away. There is no violation of well established physical laws on this view (for instance the principle of the conservation of momentum) since we have well-behaved physical particles all the way down (well maybe most of the way down – but I'll leave quantum particles to one side). Secondly, we get a much simpler view of the world – one type of phenomenon subject to one set of laws. There is no need to account for consciousness by introducing some new set of fundamental laws into our story of everything (anyone sympathetic to Occam and his razor will find something to cheer about here).

Further, as a general claim in support of the reductive account of mind, science has been incredibly successful in explaining all sorts of phenomena reductively. Things of a complex nature which used to be considered as real 'in themselves' turned out to be made of up things of a simpler nature, but in a way that held them to be just the same thing as the simpler phenomena. By way of a well worn example, it was discovered that 'temperature' is in fact 'mean molecular kinetic energy' – not, of course, that the mean molecular kinetic energy of your bath water *causes* (still less is just correlated with) its temperature. Given this success it might then strike you as odd that the process of explaining things which has been applied all the way up to biological mechanisms (think, for instance about 'life' and how it can be explained reductively in terms of complex bio-chemical processes, which themselves can be explained in terms of simpler phenomena) suddenly grinds to a halt when we reach consciousness. As J.J.C. Smart said: 'That everything be explicable in terms of physics ... except the occurrence of sensations seems to me frankly unbelievable.'<sup>7</sup>

### *The Multiple Realisability Argument*

'Well, you had better believe it', said a number of philosophers. One was Hilary Putnam in a paper called 'The Nature of Mental States', which purported to show that the type-identity theorist's project was hopeless. His argument went something like this:

If the type-identity thesis is to make good the claim that some mental property type M is identical to some physical property type P then it must provide an account of how it is that for some token<sup>8</sup> m of M, m can be realized not merely by tokens of type P but also, under different circumstances, by tokens of types P<sub>1</sub>, P<sub>2</sub> ... P<sub>n</sub> (where these are

tokens of distinct types of physical properties). According to Putnam, that is, the defender of type-identity must have it that for any organism *S* in *M* there must be some *P* had by *S* necessarily. Since some *S* could have *M* in virtue of having *P*<sub>1</sub> but lack *P* entirely then, simply, *M* can't be identical to *P*. Conversely, if the having of *P* does not guarantee the having of *M* then, once more, *M* and *P* don't appear to be identical.

Let's go a little more slowly. Suppose we take 'being in pain' as our specified type of mental property *M*, and then specify that the neural substrate (or brain property) type, *P*, with which it is supposedly identical, is some C-fibre stimulation. Now, 'do octopi have pains?' asks Putnam. He supposes so. 'Do octopi have C-fibre firings?' He supposes not. And if his suppositions are right, and they seem to be, then we have a clear counterexample to the type-identity (*M*=*P*) thesis. Identity claims being what they are this just won't do. If *M* is supposed to be identical to *P* then wherever there is a *P* there ought to be an *M*, and wherever there is an *M* there ought to be a *P*. If Cicero=Tully, then when Cicero is in the bath so is Tully, if Tully is combing his hair, then so is Cicero. In short, whatever is true of one of the 'pair' in an identity relation must be true of the other. In the case of the octopus an *M* is present but no *P* (and, for the converse, if there were a zombie<sup>9</sup> who had *P* then there would be, by stipulation, no *M*). So, either there is something wrong with Putnam's argument or we ought not to claim that *M*=*P* with respect to pain. The same will be true for any mental state that can be realized by distinct types of physical properties.

For a number of commentators Putnam's argument pretty much decided things against the type-identity thesis. Better to say, the conclusion ran, that mental states are identifiable with *functional* states which then allows for their multiple realizability in all sorts of physical organisms (perhaps even in non-physical ones, like angels, as well).

### *Why is Multiple Realizability a Problem for the Reductionist?*

Jaegwon Kim<sup>10</sup> offers an analysis of why, exactly, 'the Multiple Realization Thesis' (or MR) is standardly taken to be problematic for the identity thesis. Among the reasons given for the death of type materialism is, firstly, that if mental state types can be realized in heterogeneous ways then they cannot be *defined* physically. The idea here seems to be that any definition *D* of some mental state type *M* will be at best disjunctive since tokens of *M* are realizable by tokens of *P* or *P*<sub>1</sub> or *P*<sub>2</sub>... or *P*<sub>*n*</sub>. What this means is that the definition of 'is in pain' would include not just one neat physical property, but a rather wild, open-ended, collection of physical properties – 'has brain property of a human variety or has brain property of an octopus variety or ...' and so on. But a disjunction of such properties isn't itself a physical property (in fact, it might be said, it is a proposition and not a property at all, and no physicalist is going to claim that mental states are identical to propositions, since these are abstract entities).

Secondly, and similarly, there can be no *reduction* of a single psychological property to some basic physical property because the former property needn't involve any relation at all to the latter (certainly in some possible world there could be just such and such a psychological state without there being the physical state that, in this world, it is supposed to be reducible to). If the putative reduction base isn't the same

across all possible worlds then it's hard (well, impossible) to see how it (the absent physical base) can just be the same as the very real mental phenomenon in question.

What, then, are the consequences of MR? Well, as suggested above, we would need to characterise mental properties differently in order that they may be predicable of a range of distinct (physical) phenomena. The way Putnam thinks this should be done is to characterise them as functional properties. So, what are functional properties?

A functional property is to be understood dispositionally, or in terms of the causal role that it plays. It is easy to see how this works with something like our pain case. You and the octopus have just been badly bitten by something and we would say, ordinarily, that both of you are in pain. What this means is not that the two of you share some physical state type or property, but rather that you are both in the same functional state or possess the same functional property. Pains typically arise from bodily damage, are typically related to a system's other internal states, and typically manifest themselves in some sort of behaviour (usually behaviour related to fleeing the cause of damage, or attending to the damage, perhaps involving seeking some means of repairing it, and so on). How is it then that these very different organisms, with their wildly distinct physiology and neurology, can both be said to be in the same state, or both sharing some *one* property?

Well, says the functionalist, functional properties are 'higher level' properties. They come with lower level properties at their base, to be sure, and it seems sensible to say that the base will be a physical base, but that they are not simply identified with the base. Functional properties are had by different things insofar as these things have the appropriate dispositions given a range of inputs and internal workings, and which in turn generate a series of outputs. These functional, or higher level, properties are *realized* by a range of lower level properties. The realizers may be octopi neurological properties, or human neurological properties, or perhaps they might even be non-neurological properties (properties had by machines or Martians).<sup>11</sup>

### *What Now?*

How might one respond to MR if one were inclined to hold on to something like reductive physicalism? Here are some suggestions. Firstly, one could try to find a way of accommodating disjunctive properties in one's account – for instance showing that as second-order properties they needn't be considered as physical for a physicalist account to go through.

Secondly, one might show that mental states, globally conceived, are not individuable as kinds. On this view there would be no such thing as 'pain as such'. While this approach might appear to have an eliminativist ring to it, it doesn't actually do away with mental states such as pains,<sup>12</sup> but rather dispenses with Pain as a state applicable across a multitude of distinct physical organisms (and, perhaps, some non-physical ones too). So different pains (in Martians, in octopi, in humans) get individually classed as kinds but Pain, supposed to include all of the instances in one kind, turns out not to be a kind. So, the pain that one gets from treading on a dead bee with one's bare foot (sharp, hot, centripetal waves of it, as I discovered to my cost this past summer) still really exists and, further, there is no need for 'wild' disjunctive properties to account for it.<sup>13</sup>

On both suggested saving models mental concepts are taken to be diffuse considered in one sense, and simple considered in another. On the first account Pain (big P) turns out to be rather coarse-grained and picks out a whole cluster of things (supportably as a disjunction of properties, or insupportably as a hodgepodge of disparate properties, not really classifiable under the one concept). On the second account, pains, relativized to a subject or collection of similarly constituted subjects are perfectly acceptable as properties and neatly co-variant with the brain properties (or whatever) that they get identified with.

Some problems with the second sort of strategy spring immediately to mind. If the model is going to be acceptable then it better account for all mental states in this way (or explain why pain is a special case, and deal with the other sorts of mental states differently). So, one might ask, are properties like 'being hungry', or having some desire, or some belief or other, relativizable to species? If a Martian believes (while visiting Earth and watching the Old Trafford Test Match on TV.) that 'it is raining in Manchester' is this a case of Martian belief, distinct from human belief, or is the same belief (the same mental property) being had by two distinct sorts of organisms?

If we do relativize all mental states to species or collection of subjects (and perhaps to one and the same subject across different times and in different circumstances) aren't we in danger of multiplying perfectly intelligible states endlessly, and in a rather ad hoc fashion, simply as a means of escaping the MR claim? Well, perhaps. What seems true, at least, is that psychology, considered as a science of mentality wherever it is found, no longer looks very unified on this account. We might even struggle to neatly carve it up into sub-branches, such as 'human psychology', or 'bat psychology', or 'Martian psychology', since human pains can be realized by distinct physical types of brain properties in different persons (and in the same person at different times).

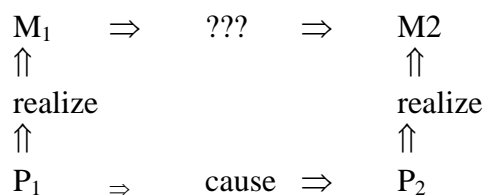
But at this point someone might ask why go to all this trouble to avoid MR. It seems perfectly natural to think of distinct sorts of physical things having one and the same property so why not just accept that higher level mental properties cannot be reduced to their lower level bases and that's that. Well, I am going to finish with a suggestion why things aren't so easy for the non-reductive view, despite the apparent reasonableness of MR.

### *Non-Reduction and Mental Causation*

I suggested earlier that one of the reasons for finding reductive physicalism attractive was that the intimacy of the identity relation provided us with the resources to solve the problem of mental causation. The thought was that if mental properties just are physical properties, and we have no problem understanding how physical properties are causally efficacious,<sup>14</sup> then mental causation comes for free.

Now, the non-reductive physicalist wants to give an account of mental causation too, keeping the mental real (i.e. not wanting to eliminate it) and making it entirely dependent on, but irreducible to, the physical. The question then arises, how do mental properties make a difference?

Suppose that we accept the non-reductive physicalist's claim that for *every* mental property there is some physical property that realizes it. Suppose also that we are happy enough with the view that for every physical effect there are physically sufficient causes for that effect. And finally, suppose we are also accepting of our non-reductive physicalist's belief that physical effects are not systematically overdetermined – that is, she takes it that, most of the time, for some particular effect, *e*, that *e* was brought about by some one cause, *c*, and that *e* would not have occurred were it not for the occurrence of *c*.<sup>15</sup> Taking all of these things together we have enough to raise a worry about the causal efficacy of the mental. Here's how:



(Figure 1)<sup>16</sup>

In figure 1  $P_1$  caused  $P_2$ , where the  $P$ s in question are physical properties, and this seems to be reasonably uncontroversial (at least we'll assume so here). Now each  $M$ , or mental property, in this story is dependent on its physical realizer, so the question is, what work is  $M_1$  doing? If we take  $M_1$  to be the desire to do something or other, and  $M_2$  the intention to bring it about that that something or other be done, we can ask how  $M_1$  is relevant to the bringing about of  $M_2$ . If  $P_1$  realizes  $M_1$ , and  $P_2$  realizes  $M_2$ , and  $P_1$  causes  $P_2$  (and hence  $M_2$ ) to come about, then is there anything at all left for  $M_1$  to do? It doesn't seem to be relevant to the bringing about of either  $M_2$  or  $P_2$  (and it's hard to see how it could be doing any work with respect to its own realizer  $P_1$ ).

We have assumed, remember, that all physical effects are covered wholly by physical causes, and that physical effects are not overdetermined, for the most part, so is there any room for the mental to enter into the causal story. If not, then we have the unfortunate consequence that the mental is *epiphenomenal*, that is to say, it is causally inert.

Why should we worry about epiphenomenalism? Well, according to some philosophers if epiphenomenalism were true then it would be wrong to say that, for instance, we speak or act at all.<sup>17</sup> Or, even more dramatically, as Jerry Fodor says, if it were true that the mental had no causal powers, then 'practically everything I believe about anything is false and it's the end of the world.'<sup>18</sup>

The argument I have sketched out (roughly) here is sometimes called the 'exclusion' argument, since it excludes mentality from the causal story. The mental appears to be indispensable to an account of the right causal story for why we do the things that we do, however. It is certainly hard to see how we could be responsible for the things we do if the causes operate at a level below our awareness.

If we want to be non-reductive physicalists then, and also keep the mental operative in the causal nexus, then we should have to find something wrong with the exclusion argument. There may well be something wrong with it, although I am not going to pursue that discussion here. I simply wanted to point out that rejecting the reductive

picture of mind presents its own set of thorny philosophical problems. As Kim says, there are no ‘metaphysical free lunches’ in this debate, so while MR seemed to push us away from reductionism, the exclusion threat might, in the end, lead us back there.

Paul Sperring

*Richmond-upon-Thames College*

---

<sup>1</sup> The mind-brain identity thesis was popularized particularly by Australian philosophers such as U.T. Place and J.J.C. Smart.

<sup>2</sup> I intend to talk primarily about mental *properties*.

<sup>3</sup> ‘[The] word ‘reductionism’ seems by now to have acquired a negative, faintly disreputable flavor – at least in the philosophy of mind. Being a reductionist is a bit like being a logical positivist or member of the Old Left – an aura of doctrinaire naïveté hangs over him.’ Kim (1989: 32)

<sup>4</sup> I take no credit for these arguments, they can be found in the works of J. Kim and others.

<sup>5</sup> ‘Types’ are to be distinguished from ‘tokens’. An itch or a pain considered in general is a mental *type* – ‘the itch that I am feeling right now in my left ear’ is a *token* of the type.

<sup>6</sup> Of course, metaphysics might have something to say about that.

<sup>7</sup> Smart (1959) ‘Sensations and Brain Processes’ quoted in Kim *Philosophy of Mind* (1998: 53)

<sup>8</sup> Type/token talk is, crudely, just general/particular talk. For example, a type of property, such as ‘being red’, is had by three red snooker balls. We have in this case one property type, and three tokens of that type.

<sup>9</sup> A zombie is a theoretical entity that is supposed to be a molecule for molecule replica of you minus all of your mental properties (well, specifically the conscious ones, such as your qualia). If zombies are possible beings then, the argument runs, mental properties cannot be physical properties. I don’t intend to say anything more about zombies here.

<sup>10</sup> ‘Multiple Realization and the Metaphysics of Reduction’

<sup>11</sup> I am not here saying anything about the different sorts of functionalist positions (as with physicalism there are many different versions of functionalism), but rather am lumping together any position that sees mental states as (a) functional states and (b) multiply realizable.

<sup>12</sup> Not that eliminativists ever *really* deny that there are such phenomena – although they are often caricatured as taking us to be insensate automata.

<sup>13</sup> Jaegwon Kim (1992) discusses strategies of these sorts, although my account of them shouldn’t be thought to be a detailed and accurate rendering of his position – I simply offer here a boiled down version of the general reductionist program that Kim offers there and elsewhere.

<sup>14</sup> Not everyone accepts that physical causation is metaphysically unproblematic – but there appears to be a long tradition of philosophers happy to accept that if only we could make sense of mental items in terms of physical items we would thereby make sense of their causal status. At the very least the ontologically simpler view on the cards with reductive physicalism means that we avoid the troubles with causation across completely distinct phenomena – which was thought to be what did for Descartes.

<sup>15</sup> This wouldn’t be the case if the effect were overdetermined, since if the cause had been missing the effect would still have come about (because the other, overdetermining, cause would have done the work). When the tightrope walker falls but is attached by means of a wire to the circus tent’s roof, and has a safety net below her, we have a case of the overdetermination of the effect of her *safely* falling.

<sup>16</sup> I take this from John Heil who raises worries about the non-reductive physicalist view in his book *From An Ontological Point of View*

<sup>17</sup> Norman Malcolm ‘The Conceivability of Mechanism’

<sup>18</sup> Fodor ‘Making Mind Matter More’ p. 156