Concessionary Dualism and Physicalism

WILLIAM SEAGER

Abstract
Modern physicalists frequently offer the generous concession that although dualism is false, it is not a metaphysical impossibility. And it appears that the proper formulation of physicalism allows for this concessionary position. It would be expected that dualists also could accept that while physicalism is false, it too is a metaphysical possibility. I will argue that a careful analysis of physicalism and dualism shows that in fact these concessionary positions cannot be maintained. In particular, the nature of the metaphysical determination relation which holds between matter and mind on both physicalist and dualist views precludes either from allowing that the other is a metaphysical possibility.

1. What is Physicalism?

The doctrine of physicalism can be roughly spelled out simply as the claim that the physical state of the world determines the *total* state of the world. However, since there are many forms of determination, a somewhat more precise characterization is needed.

One obvious problem with the simple formulation is that the traditional doctrine of epiphenomenalism holds that the mental is determined by the physical. However, the orthodox view, which seems obviously correct, is that physicalists would and should balk at the claim that epiphenomenalism is a form of physicalism.

The philosophical zombie thought experiment vividly reveals exactly why epiphenomenalism is not a version of physicalism. According to traditional epiphenomenalism the determination relation in question is causation, and causal relations do not hold with full necessity. They hold with at most nomological necessity. Thus, if epiphenomenalism is true, there is a possible world, w, in which the causal laws are different in such a way that the physical states which in the actual world cause mental states either cause no mental states at all in w (the zombie option) or cause aberrant mental states (the inverted spectrum option). Why should the physicalist care about this 'mere possibility'? Because it shows that the mental can vary independently of the physical and there can be no

doi:10.1017/S1358246110000147 ©The Royal Institute of Philosophy and the contributors 2010

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

better demonstration of ontological distinctness than independent variation.

Therefore, physicalism requires that the sort of determination at issue must exhibit *maximum* modal force; it must be absolutely impossible for the mental to vary without attendant, determining physical variation (let us label this relation 'logical determination'). I think the best way to state physicalism which meets this constraint is in terms of what are called minimal physical duplicates (MPDs) of possible worlds, an approach pioneered by David Lewis.[1] An MPD of the actual world is a possible world which is physically exactly similar to the actual world and contains nothing else. In other words, an MPD of the actual world is configured in just the way that physicalists assert the actual world is configured. Physicalism is true of world w, then, just in case w is its own MPD, or, more formally (letting 'Px' stand for 'physicalism is true at world x' and 'Mxy' for 'x is an MPD of y'),

$$\mathbf{MPD-P}. \quad (\forall w)(Pw \equiv (\forall v)(Mvw \rightarrow v = w)).$$

Obviously, the identity condition rules out the existence of *any* possible world which exhibits independent variation of non-physical features relative to the physical. So we have the desired maximum modal force. At the same time, a very nice feature of this definition is that it permits physicalism to be a contingent truth about the actual world without weakening the modal force of the determination relation. That is, even if physicalism is true of the actual world there are still other possible worlds where it is false. A world where Cartesian dualism was true would be a world where physicalism is false, but it may for all that be a genuinely possible world, and it would seem not to threaten the truth of physicalism.

The definition nonetheless does have some odd consequences. It utterly fails to even hint at the nature of the link between the fundamental physical features of the world and the higher level structures which depend on them.

One attractive way to avoid this consequence is to embrace an additional constraint on the definition of physicalism: the in principle *a priori* deducibility of the mental description (or whatever high level description is at issue) from the purely physical description, given

[1] D. Lewis 'New work for a theory of universals', *Australasian Journal of Philosophy* **61** (1983), 343–77. See also F. Jackson, *From Metaphysics to Ethics* (Oxford: Clarendon, 1998) and D. Chalmers and F. Jackson, 'Conceptual Analysis and Reductive Explanation', *Philosophical Review* **110** (2001), 315–61.

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

possession of the relevant high level concepts. Such an approach, though not exactly the one I outline here, is endorsed by both David Chalmers and Frank Jackson.[2] Very roughly, the idea is that if physicalism is true then there is a sense in which all truths in world w should be deducible from the purely physical description of w. For example, consider the set of chemical truths. These are not explicitly represented in the purely physical description of the world but – it is very plausible to believe – they are logically determined by the purely physical truths. What is more, it seems also to be true that given the purely physical description *plus* a grasp of the concepts of chemistry, the chemical description of the world could be deduced.

We might label the additional condition of *a priori* deducibility 'logical transparency' (the resulting characterization of physicalism would simply be that it is logically transparent that the defining condition of MPD-P holds). The motivating advantage of logical transparency in the definition of physicalism is that it provides the requisite link between fundamental and higher level features.

Logical transparency can also be regarded as a benchmark of the explanatory adequacy of physicalism. If physicalism is true we expect that, in principle, there *should* be an account of how or why higher level concepts apply to the world. We can hope that it is accessible to us even though the ideality condition means there is no guarantee.

Another problem with MPD-P is that it threatens to make physicalism come out trivially true under certain conditions in which physicalism is, intuitively, false. Consider a universe, call it Diffuse, exactly like ours with respect to physical law but in which the initial conditions of the big bang left it in a state where the originally created elements ended up so smoothly and sparsely distributed that stars and galaxies could never form. It is easy to believe that any MPD of Diffuse is completely identical to Diffuse – there is just nothing going on in that world except basic physical processes. So physicalism comes out true in Diffuse. But I think this is worrisome. This imaginary world seems as capable of possessing the intrinsic 'power' needed to generate biology and mentality as does the actual world.

Now, suppose that physicalism is false of the actual world. Then I think it should also be false of Diffuse as well. Could champions of physicalism make the doctrine true simply by wiping out all

[2] D. Chalmers and F. Jackson, 'Conceptual Analysis and Reductive Explanation', *op. cit.*

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

sentience? One could maintain that an MPD of Diffuse will not be identical to Diffuse since it will lack some extra non-physical laws of nature, the laws of emergence if you will, that would kick into action if and when the right kind of complex physical structure should appear (which as a matter of fact will never happen in Diffuse). But then one might worry about the status of fundamental laws (that is, laws which are not derivable from other laws) which are never instantiated in a world. And one might have Humean scruples in favor of the idea that laws supervene on regularities, and the requisite regularities will never appear in Diffuse.

There is a simpler remedy which avoids these metaphysical conundrums. It also has the advantage of explicitly recognizing that physicalism is best regarded as a doctrine about whether certain specific target phenomena are logically determined by the physical. Physicalism in general is the claim that all phenomena are logically determined by the physical, but more restricted forms are possible, and the general form is nothing more than the logical sum of all the restricted forms. The present position of strength which physicalism enjoys in philosophy stems from the relentless assimilation of ever more domains of phenomena under the purview of physicalism.

Let us then take physicalism to be the domain relative claim that for any possible world and domain of facts, d, there are some d-truths which hold at that world and all its d-truths are true at any MPD of that world, or, more formally (using 'Pdw' for 'physicalism with respect to domain d is true at w', 'p_d' for d-propositions and 'Txy' for 'x is true at world y'):

$$\textbf{MPD} - \textbf{PD} \quad (\forall w)(\forall d)(Pdw \equiv (\exists p_d)(Tp_dw) \wedge (\forall v)(Mvw \rightarrow (\forall p_d)(Tp_dw \rightarrow Tp_dv)))$$

Thus, physicalism about the mind is only going to have a chance to be true of worlds where there is mentality. Diffuse is not a world where physicalism about the mental is true (albeit for the trivial reason that there are no minds in Diffuse). A more sophisticated definition could hold that physicalism about a domain is neither true nor false in a world where that domain is not instantiated. MPD-PD has the advantage that physicalism could turn out to be true of chemistry but false of psychology or, more specifically, true of intentionality but false of phenomenality. General, or total, physicalism is the doctrine that physicalism is true for all instantiated domains. In this case, MPD-PD reduces to MPD-P given – what seems obvious – that any two possible worlds that agree on all truths are the same world.

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

2. Zombies and Physicalism

The zombie argument against physicalism about phenomenal consciousness is crystal clear. It purports to show that consciousness can vary independently of the physical. There could be no better argument for ontological distinctness.

The relation between physicalism and zombies is more interesting than this, however. They are in much stronger opposition than the above argument reveals, because if zombies are possible then physicalism is not only false, it is *necessarily* false. Here is the argument for that:

(1) $$\Diamond Z \rightarrow \neg P.$$

Here, Z = there are zombies, P = physicalism is true. To say that there are zombies is, of course, shorthand for saying there are creatures which are physically identical to targeted conscious creatures in the world of evaluation but which lack consciousness. To say that physicalism is true is shorthand for saying that physicalism about consciousness is true.

(1) follows from the argument presented above. That argument was entirely *a priori*, so we can infer that

(2) $$\Box(\Diamond Z \rightarrow \neg P).$$

One might regard this move as somewhat swift. It is not after all *a priori* true that consciousness exists. But an argument for (2) is easily provided. (2) is false if there is a possible world where zombies are possible but physicalism about consciousness is true. As noted above, physicalism about consciousness is true at a world only if consciousness exists in that world. So we need only consider a world in which consciousness is instantiated, physicalism is true and zombies are possible in that world. Obviously, such a situation is inconsistent. (2) is vindicated.

From the fact that necessity distributes over implication, we deduce from (2)

(3) $$\Box\Diamond Z \rightarrow \Box\neg P.$$

It is evident (from the appropriate modal logic for metaphysical possibility, which is S5-like) that if zombies are possible then it is necessary that they are possible, that is, $\Diamond Z \rightarrow \Box\Diamond Z$, from which we can infer the desired

(4) $$\Diamond Z \rightarrow \Box\neg P.$$

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

What should we make of this? First, if zombies are so much as possible then there are no possible worlds where physicalism about consciousness is true.

Therefore anyone who thinks that zombies are possible is going to 'know' that physicalism *cannot* be true. The argument is not very complex, so quite minimal rational reflection on the part of someone who regards zombies as possible reveals that the truth of physicalism is inconceivable. It is inconceivable in the same sense that the idea that there is a greatest prime number is inconceivable – rational reflection shows that the concept of prime number is inconsistent with there being a largest prime number. So, it is genuinely inconceivable that physicalism is true given a commitment to the possibility of zombies.

3. Stalemate?

By applying *modus tollens* and some simple modal-negation manipulation we can derive the *physicalist's* version of (4):

(5) $$\Diamond P \rightarrow \Box \neg Z.$$

This shows that physicalism and the possibility of zombies are 'modal contraries', and the mere possibility of one entails the logical impossibility and hence inconceivability of the other. I think this does reflect the kind of 'rock bottom' level of disagreement between defenders of physicalism and those who defend some kind of dualism. Perhaps this is an irresolvable clash of intuitions.

But is the dialectical situation really one of intellectual stalemate here? I think this is unlikely to be the case and that the dualists have a distinct advantage. The argument would stall completely if the intuition in favor of the possibility of zombies and in favor of physicalism were equally strong. This does not seem to me to be the case, however. There is no basic intuition in favor of physicalism. Its strength springs from the history of success which science has enjoyed over the last 400 years or so, as more and more of the natural world has been linked to the entities and processes described by fundamental physics. Although this linkage is undeniably supportive of physicalism, it has always been incompletely specified and seems sure to remain so, if for no other reason than the staggering immensity of the details of the determination relation which holds between the fundamental microphysical world and the macrophysical world of the 'manifest image'. And it is not simply details that stand in our way; we have at present no good understanding at

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

all of how the observed world can emerge from the underlying quantum reality.

On the other side of the coin, the possibility of zombies is a clear extension of age old worries about the highly significant gulf between what we can observe of people and their mental states. As Shakespeare expressed it in *Macbeth*: 'There is no art to find the mind's construction in the face'.[3] The idea of the mindless automaton which counterfeits human action is also very familiar, as is the notion of a trance state which renders someone unconscious but capable of action. The philosophical problem of other minds is yet another familiar difficulty, which can be explained to non-philosophers in just a few minutes, that reinforces the apparent possibility of zombies.

The zombie conception goes much further than these considerations of course, but it is worth remembering that this thought experiment is simply the extreme end of a continuum, any point of which will refute physicalism. Instead of a zombie, we can imagine a possible world physically identical to this one but in which the precise hues of experienced colors are ever so slightly different than in the actual world. This is, on the supposition of physicalism, absolutely impossible but it is hard to see exactly what it is about the physical structures which so perfectly and definitely constrains all the details of the mental with the inexorable force of logical necessity.

It may be that future science will reveal how the existence and precise qualitative features of subjective experience are logically determined by physical states and purely physical laws. It was, after all, ignorance of the laws of physics that permitted philosophers from J. S. Mill to C. Lloyd Morgan and C. D. Broad[4] to deny that chemical properties were necessitated by underlying physical structure. But since ignorance (leavened with rational reflection) is a legitimate ground for at least tentative judgments of possibility, it seems to me that the hypothesis that zombies are logically possible, despite its philosophically extreme nature, is less epistemically extreme than the physicalist hypothesis. The basis of the former is a kind of (putative) insight into the nature of consciousness, the latter is based upon an inductive inference from an incomplete and possibly irrelevant database.

[3] Act 1, Sc. 4.

[4] J. S. Mill, *A System of Logic, The Collected Works of John Stuart Mill* (Toronto: University of Toronto Press, 1963), Vols. 7–8; C. Morgan, *Emergent Evolution* (London: Williams and Norgate, 1923); C. D. Broad, *Mind and Its Place in Nature* (London: Routledge and Kegan Paul, 1925).

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

4. Concessionary Physicalism and Dualism

Perhaps reflections such as the above are what prompt the concessionary spirit in physicalists, who allow that while physicalism is true, dualism is possible.[5] Recall that the definition of physicalism allows that there could be possible worlds in which Cartesian dualism provides the true account of the mind-body relation. Such worlds are logically compatible with the idea that the physical logically determines the mental so long as none of those worlds are physically identical to the actual world.

It would be natural to think that there should be an argumentative symmetry here. That is, it might be expected that just as the physicalist will concede that there are possible worlds where physicalism fails, so too the dualist would also allow that there are worlds where physicalism is true. If so, then following the argument given above it turns out that according even to the dualist zombies are impossible.

That the possibility of zombies does not logically follow from the acceptance of dualism is quite interesting. But while this may be strictly correct, it seems very uncomfortable to hold to both the possibility of dualism and the impossibility of zombies. Consider a possible variant of a standard Cartesian world, a world that is physically just like Diffuse but in which Cartesian immaterial minds exist, though they are not anchored to any physical entity (perhaps they are angels). Physicalism is clearly false at such a world because, while mentality is instantiated in this world, its MPDs will lack all mental features. There are no zombies in an MPD of Cartesian-Diffuse, but if *all* minds were Cartesian minds the path toward the possibility of zombies is not hard to make out.

Thus, the concessionary dualist who wishes to allow for the possibility of physicalism is going to have to say that while some minds are – or may be – Cartesian, others are physical. Unfortunately, it is not very clear what a dualist could mean by this.

In line with our understanding of physicalism, dualism should be conceived as a claim about the actual world: the mental features of the actual world are radically non-physical; physicalism is false of the actual world. But, I will argue, on this reading, any dualist should be strongly inclined to think that zombies are possible.

For consider that according to the concessionary dualist the possibility of zombies is prevented only by the bizarre modal fact that it is

[5] See for example F. Jackson's 'Finding the Mind in the Natural World' in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan and Güven Güzeldere (Cambridge, MA: MIT Press, 1997), 483–92.

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

somehow impossible for a putative MPD of the actual world to lack mental features but retain a physical identity with the actual world. But according to the dictates of MPD physicalism, that would mean that physicalism would be *true* of the actual world, not false. Dualism would not stand in opposition to physicalism in its claims about the way things actually are.

The problem is with the interpretation of the Cartesian idea that without input from the mind, physical stuff could not (and this must be a *logically* 'could not') act as it does in the actual, enminded world. Such a modal constraint is very hard to grasp. In any case, this is an incredibly thin reed to hang a doctrine upon.

In the first place, it is hard to see how the dualist could avoid the possibility of 'partial zombies'. These are creatures physically identical to a Cartesian enminded creature during a period of time in which there is no interaction between body and mind (perhaps during a period of purely intellectual reflection). Given the appropriate construction, one could define a suitable partial MPD: a short lived possible world physically identical to the actual world during the period of zero interaction but lacking all Cartesian minds. Such a world would serve to refute physicalism via the possibility of zombies. The dualist can disallow such a world only if it is logically impossible for there to be a period of time during which mind and body fail to interact.

Furthermore, a purely physical world in which there was indeterminacy (such as quantum mechanics suggests obtains in the actual world) would seem to be able to present a purely physical appearance in complete accord with that of the actual world, even if Cartesian minds are at work there. Wherever in the actual world a mind stepped in to effect a change in physical processes, let a simple jolt of indeterminacy direct the material process in the same way, though without the aid of any mental cause.[6] Arguably, such a world is an MPD of the actual world, but one without any mental features. In such a world, there are zombies.

[6] Karl Popper and John Eccles suggested in *The Self and Its Brain* (New York: Springer International, 1977) that the mind could exploit quantum indeterminacy to operate in the physical world, neural synapses being the 'seat of the soul', with no apparent violation of physical law. It follows that there could be an MPD of an Eccles-Popper world that lacked mentality; that is, that contained zombies. Of course, under the right choice of measurement sequences the Eccles-Popper world would look extremely unlikely relative to the calculated quantum probabilities of synaptic transmission, but extremely unlikely is a long, long way from impossible.

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

Perhaps the concessionary dualist could object that in the actual world there is no indeterminacy. Despite appearances the physical laws are all completely deterministic (but are broken by mental intervention upon occasion), and so our scenario fails to yield an MPD of the actual world. This is the same thin logical reed. It would ultimately entail that according to the concessionary dualist it would somehow be logically impossible to construct an MPD of the actual world without including non-physical mental features in it. The postulation of such a bizarre, brute modal fact does not seem to count exactly in favor of dualism.

One potential way for the dualist to support this bizarre necessity is to endorse a causal essentialist account of property identity. The dualist can then explain why there are no MPDs of the actual world which lack mental properties without giving up dualism. For it is then open for the dualist to claim that it is strictly necessary that any MPD of the actual world will instantiate mental properties, since it will be an essential characteristic of the physical features of the world (when properly arranged) that they causally generate mental features.[7] There are many difficulties with this response. One is that even if we endorsed causal essentialism for properties it would not follow that *all* causal powers of a property are essential to it. It might be that the essence of physical properties lies entirely in their power to interact with purely physical aspects of the world (even if they also possess contingent powers to generate non-physical features). A more serious difficulty stems from the option available to the physicalist to refine the definition of minimal physical duplication, restricting the causal powers of physical properties to physical effects. This seems reasonable given that physical causation resolves itself into the fundamental physical features of the world (see section 5 below). If it turns out that mentality will not arise in a world whose physical properties have been thus restricted then physicalism has been falsified even if there is a sense in which the restriction means that we have altered the identity of the physical properties which inhabit the actual world. The dualist must at this point simply insist that this restriction is somehow impossible – that it is inherent in matter that it causes non-physical effects.[8] This does seem

[7] This idea is deployed by Brian Garrett in the context of an anti-zombie argument. See his 'Causal Essentialism versus the Zombie Worlds', *Canadian Journal of Philosophy* (forthcoming).

[8] Such a dualist would be endorsing a form of what is often called Russellian Monism, but not one which could be allowed to count as a form of physicalism. See D. Chalmers, *The Character of Consciousness*

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

extremely implausible and thus it seems to me that a dualist ought to embrace the possibility of zombies, and hence deny the possibility – and hence the conceivability – of physicalism.

Another interesting problem then looms. The physicalist made much of the magnanimous concession to the dualist that there might be Cartesian worlds. But it seems just a small step from those possible worlds to the possibility of zombies. Now it is the *physicalist* concession which can stand only if the same thin logical reed is put in place. Cartesian worlds are possible only if it is impossible that such worlds have MPDs which lack mentality. But if Cartesian worlds are possible they will – barring bizarre brute necessities – have such MPDs. And the Cartesian world which is physically identical to the actual would will have an MPD which is an MPD of the actual world, thus refuting physicalism via the reinstatement of the possibility of zombies.

The maximum physicalist concession perhaps should be along the following lines. There are worlds just like the actual world but which include non-physical minds. Physicalism is false at such worlds since their MPDs lack the immaterial minds. But these worlds retain the mental features necessitated by the physical processes going on in them. There are no worlds in which the mental states 'associated' with physical bodies all belong to Cartesian minds which are linked to those bodies according to Descartes' picture of the substantial union. Somehow, these latter are not permissible dualist worlds.

The maximum dualist concession is … what? If dualism requires belief in the possibility of zombies, then the dualist is severely constrained. So the key issue remains whether dualism entails the possibility of zombies. As we saw above, on the face of it, strictly speaking there is perhaps no such entailment. But a closer look suggested otherwise. This because of the general nature of the mind-body relation that dualism must posit. Let us have a further look at this relation.

The general issue is the familiar one of property dependence or supervenience. A superficial survey presents a number of distinct sorts of dependence that are worth distinguishing.

(Oxford: Oxford University Press, forthcoming) for a discussion of the role of Russellian Monism in the general conceivability argument against physicalism (a version of the argument is available online at http://consc.net/papers/2dargument.html).

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

5. Dependencies

The main question at issue here is what it takes for an object to have a property. This is intended in the 'proximal' sense. Obviously, there can be long chains of causation that lead to an object's coming to have a certain property. But the issue here is what is the immediate requirement for an object to have a property. I am interested here in properties which have a dependency of some sort on other properties. Presumably there are also 'basic' properties which objects have primitively or directly and presumably the most fundamental properties are among (or perhaps exhaust) the set of basic properties.

Dependency Type 1. An object, a, can have property F in virtue of other, completely distinct, objects possessing other properties and standing in certain relations. For example, something's having the property of being Canadian currency depends upon other individuals having certain properties. Properties exhibiting Type 1 dependency are extrinsic or relational. We can roughly formalize this, leaving various niceties[9] aside, as:

$$\textbf{DT1} \quad Fa \rightarrow (\exists\Gamma)(\exists b_1, b_2, \ldots, b_n)(\Gamma b_1, b_2, \ldots, b_n \wedge \Box_l$$
$$(\forall x_1, x_2, \ldots, x_n, y)(\Gamma x_1, x_2, \ldots, x_n \rightarrow Fy))$$

where Γ stands for the properties/relations which underlie the property in question and $\Box_l$ represents full-on logical necessity. Type 1 dependency as defined is extreme insofar as it ignores any contribution of the object, a, itself. The property of 'being accompanied in the universe' would seem to be an example of pure Type 1 dependency (which immediately generates infinitely many others of essentially the same ilk, *e.g.* being accompanied in the universe with a red thing, *etc.*). Impure forms represent combinations of the types of dependencies listed here.

Dependency Type 2. Fa holds in virtue of a possessing some other property, G. We can call such properties derivative. All determinable properties are derivative from the determinate possessed (*e.g.* the property of being colored holds in virtue of the object having some definite color). Logical 'inclusion' more generally provides other examples of derivative properties (*e.g.* the property of weighing at least 9 kg derives from the property of weighing 10 kg, given that the object in question weighs exactly 10 kg). We might try to

[9] For example, the explicit clause asserting that each b is different from a. In addition, here and below I ignore outer universal quantification for simplicity of presentation.

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

formalize the idea of F being derivatively dependent on G thus:

DT2 $Fa \rightarrow (\exists G)(G \neq F \wedge Ga \wedge \Box_l(\forall x)(Gx \rightarrow Fx))$

Dependency Type 3. Fa holds in virtue of the constituent parts of a having distinct properties and participating in certain relations. Call such properties 'logically emergent'. It is crucial to note that in this sort of dependency nothing is required for a to possess F other than for its parts to have the properties in question, where these parts are evolving and interacting according to the laws applicable to those parts. For example, it suffices for something to have the property of being transparent if its constituent molecules have certain atomic/electronic properties subject to the actual laws of QED. But, obviously, if the relevant properties are hypothetically imagined to be connected in some other way than they are in the actual world, then a may or may not possess F. In a world where the laws of QED were different, perhaps transparency would not emerge. We can, more or less, formalize this notion as follows:

DT3 $Fa \rightarrow (\exists \Gamma)(\exists z_1, z_2, \ldots, z_n)(Cz_1, z_2, \ldots, z_n a \wedge$

$\Gamma z_1, z_2, \ldots, z_n \wedge \Box_l(\forall x_1, x_2, \ldots, x_n, y)$

$(Cz_1, z_2, \ldots, z_n y \wedge \Gamma z_1, z_2, \ldots, z_n \rightarrow F_y))$

where 'Cxyz' represents 'x and y constitute z'.

I think Type 3 dependency is the core idea which funds MPD based physicalism. According to the physicalist, all it takes to get mentality into the world is for the physical to be arranged thus-and-so and for the physical to operate according to the actual physical laws. Officially, however, MPD physicalism is not committed to any claims about the constituent structure of objects possessing mental properties. Officially, all that is required is that there can be no change in mental properties without a change in a physical property *somewhere* in the world at issue. That is, the physical-to-mental relation could be a Type 1 dependency (a kind of pure externalism of the mental which is highly implausible) or an amalgam of Type 1 and Type 3 dependency (something like this is what naturalist externalists about the mental have in mind). The intrusion of Type 1 dependency leads to bizarre modal structures in which there are possible worlds that contain zombies but differ physically from the actual world only in the position of a neutrino on the other side of the universe. I suggested that logical transparency would eliminate such modal oddities since the position of the neutrino presumably

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

would not figure in the deduction of mental properties from the physical ground. We could thus absolutely rule out that the 'shifted neutrino' world is a zombie world.

Dependency Type 4. Fa holds in virtue of the constituents of a having exactly the same property. Call such properties compositional. Mass is an example and, if we take specific mass properties, is actually highly interesting. For example, the specific mass property exemplified by a hydrogen atom is 1.007825037 amu. The mass of a proton is 1.00727638 amu and that of the electron is 0.000548579867 amu. The atomic mass is not quite the same as the sum of the masses of the components because the energy which binds the electron to the proton must be taken into account according to the relativistic principle of energy-mass equivalence. This latter is an empirical law which governs the causal interaction of the proton and the electron. The empirical principle involved could have been different. But, crucially, the principle is a physical law which operates over purely physical properties.

There are many other examples of compositional Type 4 dependency. However, all the examples that come readily to mind of true compositional properties are physically fundamental. Massive things get their mass from the mass of their components, charged things get their charge from the charges of their components, *etc*. It is hard to think of genuine cases of non-fundamental compositional properties. Initially appealing counterexamples to the claim that compositional properties are fundamental arise from almost any 'mass term'. Water is made of water, gold of gold, and so on. However, all such substances resolve themselves into more fundamental structures upon which the identity of the substance supervenes. This observation goes some way in explaining why it is that compositional properties seem to be fundamental.

While Type 4 dependency is, of course, a sub-class of Type 3 and needs no separate formalization, the tendency toward fundamentality of compositionally dependent properties makes Type 4 special.

Dependency Type 5. Fa holds in virtue of its parts having properties or entering relations which *proximally cause* a to have F. Call these properties 'causally emergent'. A possible formalization requires only a typographically minor alteration to that of Type 3 dependency, namely:

$$\textbf{DT5}\quad Fa \rightarrow (\exists\Gamma)(\exists z_1, z_2, \ldots, z_n)(Cz_1, z_2, \ldots, z_n a$$

$$\wedge\ \Gamma z_1, z_2, \ldots, z_n \wedge \Box_n(\forall x_1, x_2, \ldots, x_n, y)$$

$$(Cx_1, x_2, \ldots, x_n, y \wedge \Gamma x_1, x_2, \ldots, x_n \rightarrow F_y))$$

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

where $\Box_n$ represents nomological or causal necessity. It is important to emphasize that this is supposed to represent the proximal ground for a's having property F (there is no shortage of 'long range' Type 5 dependency).

Although this conception seems to be perfectly consistent and it does not seem impossible that certain properties exemplify Type 5 dependency,[10] there is a big question about whether there actually are any examples of Type 5 dependency as opposed to Type 1–4 dependency. In terms of a theological metaphor, cases of Type 5 dependency would require God to create the stuff forming the constituents of an object (along with the laws governing these constituents) but also and in addition create a new fundamental law which caused the possession of property F when the constituents got into certain arrangements.

One intriguing and often suggested possibility is that quantum entanglement provides a genuine example of Type 5 dependency.[11] It is a fundamental feature of quantum mechanics that any two states of a system can be 'added' to form a new state (this is the linear superposition principle). Given certain other core features of quantum mechanics, this leads to the following possibility. Consider a system which creates particles in pairs with some conserved property, such as spin, which can take either a positive or negative value (for simplicity we can ignore the magnitude and just label these + and −). Assuming we start with zero total spin, then the spins

[10] One philosopher who has argued against the coherence of Type 5 dependence is Galen Strawson who, in 'Realistic Monism: Why Physicalism Entails Panpsychism' (*Consciousness and its Place in Nature: Does Physicalism Entail Panpsychism?*, edited by A. Freeman (Exeter: Imprint Academic, 2006), 3–31), takes the impossibility of Type 5 dependence to rule out any form of genuine emergence of consciousness from a purely physical substrate and thus to support panpsychism (which is, in my view, best understood as a form of Type 4 dependence, about which see below).

[11] Paul Teller can perhaps be read as advocating such a view, although it may be that he is only endorsing the weaker claim that quantum systems exhibit a kind of irreducible ontological holism (see his 'Relational Holism and Quantum Mechanics', *British Journal for the Philosophy of Science* **37** (1986), 71–81). I think that discussions of emergence in quantum mechanics tend to miss the distinction between holism and Type 5 dependency. This is often coupled with the assimilation of Type 3 dependency to 'part whole reductionism'. But while Type 3 dependence is compatible with mereological reductionism it is not equivalent to it, as the example of mass illustrates.

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

of the individual particles (call them A and B) must sum to zero. There are two ways that this can happen, namely, if A has spin + and B has spin −, or the reverse. There is no way to control the polarity of a particular particle's spin during pair production, so when, for example, A is created it is in a superposition of + and − spin, and similarly for B. But since the total spin of the system has to be zero, if we measure A and find it has spin + then we immediately know that B has spin − (or at least will, if measured, give a guaranteed result of −).

This state, known as the singlet state, can be expressed in this highly non-standard but hopefully perspicuous form:

$$\text{Singlet } \frac{1}{2}(A^{+} \otimes B^{-}) + \frac{1}{2}(A^{-} \otimes B^{+})$$

where the ⊗ symbol represents the 'joint state' of particle A and B. There is no way to decompose this complex superposition into a form in which the A states and B states are separated, hence the use of the term 'entanglement' to describe such states.

Entanglement has many peculiarities. Notoriously, since measurement in effect forces Singlet into one of its terms, it entails that no matter how far apart A and B are, upon measuring A to be + (−), it is instantaneously fixed that a measurement of B will give − (+). Also, it seems clear that quantum mechanics is endorsing, or perhaps revealing, some kind of holism about entangled states; they are not reducible to purely local states of the component particles. Furthermore, there is no way to devise any local properties of A and B which can 'carry' the observed correlations between possible spin measurements (though this is possible if the properties can exchange information instantaneously across any distance). So entanglement is very weird.

However, our question is about whether this phenomenon gives any support to the claim that modern science provides examples of Type 5 dependence. The answer seems clearly to be 'no'. The superposition principle and the formation of joint states are fundamental principles of basic quantum physics. Singlet is a state fully described by the fundamental laws of quantum mechanics. Just as the complexities of mass composition depend logically upon the laws governing mass and energy (which are basic physical laws), so too the complexities of entanglement are a logical consequence of the laws governing the basic properties and interactions of the quantum world. In some ways, the analogy is quite close. One would be severely misguided to think that one could simply add up the masses of the constituents of an object, considered independently, to compute the total mass

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

of the object. The mass of an object is, in a sense, not reducible to the mass properties of the individual components. Their interaction has to be taken into account. But this does not show that an object's mass has a Type 5 dependency on the mass of its constituents, because the interactions are governed by the fundamental laws at the level of the constituents themselves. Similarly, the entangled state is a predictable logical consequence of the basic laws of the quantum particles and their interactions. In fact, it *was* predicted, by Schrödinger, who introduced the term 'entanglement', at the birth of quantum mechanics. There is no hint of Type 5 dependency here, although the oddity of entanglement is also emphasized by this analogy – unlike in the case of mass, there is apparently no (current ongoing) interaction between the entangled particles!

Although modern science does not accept the existence of Type 5 dependence, it nonetheless does not seem to be impossible and has been seriously endorsed. A recent example is that of 19th-20th century emergentism. Once we allow for the possibility of causal emergence, then the step from the constituting microstructure and interactions to the macrophysical property might actually fail to fall under our Type 3 dependency relation. Many of the so-called British emergentists mentioned above believed this, and used chemistry as the lynchpin and, as it seemed to them, utterly uncontroversial example of a sort of emergence which could not be predicted nor was logically determined by sub-chemical physical processes. On their view, chemical, and many other including mental, properties stand precisely in a Type 5 rather than Type 3 dependency relation to the underlying physical structures and interactions. If the causal emergence of chemistry was actually true, there would be possible worlds where the underlying physics was the same as in our world but in which chemistry was different (or absent). Such chemical-zombie worlds are obviously an analogue to the mental zombie worlds under discussion here. They arise from the fact that basic laws of nature are not necessary but rather vary from possible world to possible world (non-basic laws are relatively necessary since they follow logically from basic laws.)

Following these precedents, a dualist of a certain emergentist bent will insist on Type 5 dependency of the mind on the physical, that there are fundamental non-physical laws of emergence which regulate the generation of mental features when certain physical configurations come into being.

Is there really a difference between Type 3 and Type 5 dependence? Type 3 is intended to express the idea that there is a logically necessary connection between F and Γ (using the latter symbol, as

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

above, as an abbreviation for the complex of properties and relations which underlie F). It is thus an intrinsically stronger claim than the one made by Type 5 dependency. If the connection described in DT3 holds between F and Γ, then there are no possible worlds where something has constituents with properties/relations Γ but lacks F. But in the case of the connection described by Type 5 dependency, all it takes to break the connection is a breakdown in causal laws, and all causal laws are non-necessary. There are possible worlds where something has constituents which stand in the appropriate relation but lacks F if the connection is only of Type 5.

Now, consider once again the dualist who – in the concessionary spirit of generous moderation – allows that physicalism might be true, that there are possible worlds in which physicalism is true. Physicalism asserts that the relation between mental properties and the physical is of Type 3.

In light of this we now face once again the question of how the concessionary dualist should understand the nature of dualism. If it is accepted that physicalism is true at some worlds, then at those worlds there is a Type 3 dependence of the mental on the physical. The dualist might think to embrace this possibility by postulating worlds in which there are non-physical structures which similarly support a Type 3 dependency relation of the mental. This is akin to the frequently advanced idea that a functionalist about the mind is indifferent about the 'stuff' which realizes mental properties. In Hilary Putnam's words, 'strictly speaking, a Turing Machine need not even be a physical system'.[12] As Putnam develops this idea, he skates over the delicate question of what would be the nature of the realizing states of a 'Cartesian mind' which was a Turing Machine. But it is obvious that if the realizing stuff is not intrinsically mental then we do not have a form of dualism in the relevant sense. In a world where a non-physical and non-mental substrate provides the ground of a Type 3 dependency, the mental is not a fundamental part of that world's ontology. This is a dualism which divides the physical from the unknown non-physical *and* non-mental substrate which realizes the mental, not the desired dualism of mind and body.

Now, it is possible to imagine a dualist who accepted both that the realizing substrate and the realized properties were mental, while also allowing that there could be a Type 3 dependence from the physical to the mental. But in such a case the mental nature of the realizing

[12] H. Putnam, 'The Mental Life of Some Machines' in *Mind, Language and Reality: Philosophical Papers, Vol. 2* (Cambridge: Cambridge University Press, 1975), 412.

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

substrate would be irrelevant to the realized mental properties. What would be relevant are the causal relations amongst the structures of the realizing substrate which can be perfectly mimicked by the structure of physical states in possible worlds where physicalism is true, and these, by hypothesis, do not depend in any way on the substrate being mental. Such a view would be a bizarre form of panpsychism, but one in which the mental nature of the subvening components would be irrelevant to the mental nature of the macro-mental features which supervene upon them. Once again, we would not have a mind-body dualism but a dualism of the physical and 'causal structure' (already familiar from standard functionalism).

More sensibly, a dualist could maintain that the mental nature of the constituents was essential for the generation of the macro-mental features. That is, such a dualist would hold that the dependency relation between the subvening constituents and supervening mental states was of Type 4 (a sub-class of Type 3). Thus it could be held that mental properties are 'composed' out of the mental properties of an object's constituents in a way analogous to the composition of the mass of an object from the masses of its constituents. As noted above, the dependency relation involved with mass is not a simple addition but involves empirical laws with some definite and complex content, albeit laws which operate at the level of the constituents. Similarly, for the dualist, principles operating at the level of the 'micro-mentality' of the object's constituents would generate its 'macro-mental' state.

Our concessionary dualist has been forced to recognize that mental features cannot have a Type 3 dependence on non-mental properties (or non-mental aspects of notionally mental properties). What about the other dependency types? Type 4 is a variant of Type 3, but if we are not allowed mental properties at the base, Type 4 cannot solve the problem. It is evident that mentality is not a logical derivative of being physical in the way required for Type 2 dependency.[13] In any case, as noted above, Type 2 dependency is, from the point of view of the physicalist, merely a disguised form of Type 3 dependency since the macrophysical property required for a Type 2 dependence would itself be dependent on the microphysical via a Type 3

[13] Stephen Yablo has offered what I take to be a surprisingly fruitful *analogy* between the mind-matter relation and the determinable-determinate relation ('Mental Causation', *The Philosophical Review* **101** (1992), 245–80). There are problems with trying to take the idea literally, however (for some, see E. Cox, 'Crimson Brain, Red Mind: Yablo on Mental Causation', *Dialectica* **62:1** (2008), 77–99).

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

relation. Nor does the mental seem to be an extrinsic or relational property, still less a relational property dependent on the non-mental. Worse, as noted, this threatens to let zombies in by the back door so to speak. If mentality is extrinsic then zombies could exist in a world with a few atoms moved just a few nanometres from their actual positions.[14] Another obvious worry here is that if the physicalist can specify the region required to subvene the mental, then the state of this region ought to underpin a Type 3 determination of the mental within that region. This leaves only Type 5 for the dualist to embrace. But Type 5 dependency of the mental on the physical leads directly to the possibility of zombies.[15]

If this is what dualism involves, then it turns out that the concessionary physicalist is *also* in trouble. If there is a possible world where dualism is true then there is a world, a different world which is specified by the dualist relative to the dualist world allowed by the physicalist, which is an MPD of the actual world but which lacks mentality. This the physicalist cannot allow. So there can be no concessionary physicalists after all.

If one is to be a physicalist, one must regard dualism as absolutely impossible (hence inconceivable). If one is to be a dualist, one must regard physicalism as absolutely impossible (hence inconceivable). For a physicalist, it seems there is no coherent notion of the non-physical, save for the merely verbal concession that some nominally non-physical 'stuff' could mimic the causal relationships enjoyed by the physical which are themselves sufficient to logically determine mentality. The dualist intuition is that there is a coherent idea of a different realm of being: consciousness, which is not logically dependent on the physical (even if it is as a matter of fact causally dependent on the physical). Though this gives every appearance of

[14] Leaving aside the 'shifted particle' worlds, it is an explicitly noted, if unfortunate, consequence of the view that mental properties are extrinsic that a creature that failed to stand in the requisite relations would thereby qualify as a zombie, no matter how physically (hence behaviorally) similar it might be to one of us; see F. Dretske, *Naturalizing the Mind*, (Cambridge, MA: MIT Press, 1995). This seems almost a *reductio* of such accounts of the mental, with the added 'bonus' of strongly supporting the general intuition of the possibility of zombies.

[15] Of course, there remain a host of dualist views in which the mental bears no relation of determination upon the physical (pretty much all the usual suspects: occasionalism, parallelism, *etc*.). But this sort of dualist will have no reason to deny that there are worlds which are MPDs of the actual world that lack mentality. Zombies remain possible.

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press

both conceivability and possibility, the physicalist dare not make such a concession.[16]

University of Toronto

[16] This paper began as a commentary on Alexandru Manafu's presentation at the 2008 CPA meeting and I would like to thank Alexandru for sparking my reflections on this topic. Additional comments of David Chalmers have been very helpful, though of such a generous extent that they could not all be addressed here.

https://doi.org/10.1017/S1358246110000147 Published online by Cambridge University Press