

# Can Physicalism Be Non-Reductive?

Andrew Melnyk\*

University of Missouri

---

## Abstract

Can physicalism (or materialism) be non-reductive? I provide an opinionated survey of the debate on this question. I suggest that attempts to formulate non-reductive physicalism by appeal to claims of event identity, supervenience, or realization have produced doctrines that fail either to be physicalist or to be non-reductive. Then I treat in more detail a recent attempt to formulate non-reductive physicalism by Derk Pereboom, but argue that it fares no better.

---

## 1. Introduction

Physicalism about the mind (henceforth, just *physicalism*<sup>1</sup>) is the claim that mental phenomena are in some sense nothing over and above physical phenomena; and obviously a central question in the philosophy of mind is and has long been whether physicalism is true. But there's also a question about physicalism that's prior to the question of whether it's true, namely, the question of how it should be formulated. This prior question raises two sub-questions. What should 'physical phenomena' be taken to be? And how must mental phenomena be related to physical phenomena for mental phenomena to count as 'nothing over and above' physical phenomena?<sup>2</sup>

But it also raises a third sub-question, which forms the topic of the present article. The question is whether physicalism, properly formulated, is *non-reductive*. One view is that physicalism, properly formulated, is *reductive*; that is, it's inconsistent with the irreducibility (i.e., it entails the reducibility) of mental phenomena to physical phenomena. A second view is that physicalism, properly formulated, is *weakly non-reductive*; that is, it's consistent with the irreducibility of some or all mental phenomena to physical phenomena, but doesn't entail it. A third view is that physicalism, properly formulated, is *strongly non-reductive*; that is, it's not just consistent with the irreducibility of some or all mental phenomena to physical phenomena, but actually entails it. Obviously, further elaboration of these views requires some account of reducibility; see below.

I myself incline strongly to the first view – that properly formulated physicalism is reductive, and hence that, to echo Jaegwon Kim's provocative language from 1989, non-reductive physicalism is (still) a myth (see Kim,

'Myth of Non-Reductive Materialism'). But I have no knock-down argument for this view. What moves me is rather that, as I see it, the history of attempts to formulate versions of (weakly or strongly) non-reductive physicalism is a history of failure. In what follows, I'll first narrate this history, and then – I hope – add to it by arguing that a recent attempt (by Derk Pereboom) to formulate a strongly non-reductive version of physicalism is no more successful than its predecessors. Through the history of failure that I claim to find there runs a very simple thread. A successful formulation of non-reductive physicalism must, obviously, be non-reductive (either weakly or strongly); but it must also be authentically physicalist. However, extant attempts to formulate non-reductive physicalism haven't managed to be both. Or so, at any rate, I'll argue.

## 2. The Background

Attempts to formulate versions of non-reductive physicalism were a reaction to the alleged deficiencies of an earlier version of physicalism that was regarded as reductive. According to this earlier version of physicalism, every type of mental state is *identical* with (i.e., one and the same as) some type of physical – presumably neurophysiological – state. On this view, then, a backache is a brain state of a certain type, thinking it's time for lunch is a brain state of some second type, wanting to take a bath is a brain state of some third type, and so on for all the types of mental state there are. Classic mid-twentieth century defenses of physicalism (e.g., Smart) seem to have assumed that physicalism should be formulated in this way, as an assertion of the identity of mental types with neurophysiological types, even though at that time little attention was paid to the niceties of formulating physicalism.

Type-identity physicalism, as we may call it, was viewed as reductive because it counted as reductive given the view of reducibility that philosophy of mind in the late 1960s took over from contemporary philosophy of science (see Sklar *qtd.* in Fodor 121–2). According to this view of reducibility, which from now on, for reasons to be made clear soon, I'll call the *neo-Nagelian view* of reducibility, reducibility, at least in the first instance, isn't a relation between two classes of phenomena (e.g., mental phenomena and physical phenomena), but rather a relation between two theories (or between two families of theories), where a theory is assumed to be a set of law-statements stated in a distinctive vocabulary. On the neo-Nagelian view, to put it roughly, one theory is reducible to another just in case the first theory can be exhibited as a special case of the second. More precisely, a theory T (e.g., a psychological theory) is reducible to a theory  $T_b$  (e.g., a physical theory) iff the law-statements that make up T are *deducible* from the conjunction of the law-statements that make up  $T_b$  with true identity-statements claiming the identity of every type mentioned in T with some type or other mentioned in  $T_b$ . Now type-identity

physicalism asserts precisely that every mental type can be identified with some neurophysiological type. So, given the neo-Nagelian view of reducibility, type-identity physicalism entails that everyday discourse about mental states (construed either as the point of departure for a scientific psychological theory or as itself an honorary scientific theory) is reducible to neurophysiology.

Moreover, the neo-Nagelian view of reducibility wasn't adopted as an arbitrary stipulation about the word, 'reducible'. It was, of course, a modified form of the view of reducibility explicitly developed by Ernest Nagel (336–66) as articulating an attractive intuition about reduction: that reduction is a species of explanation – to a first approximation, the synchronic, non-causal explanation of the behavior of entities in one theoretical domain entirely in terms of the behavior of entities in another theoretical domain.<sup>3</sup> The neo-Nagelian view articulates this intuition by entailing it. If one theory is reducible in the neo-Nagelian sense to another theory, then, given the requisite type-identity statements, the first theory is deducible from the second theory; and, in the case of a theory (i.e., set of law-statements), deducibility from another theory suffices (more or less) for its explainability by that other theory.<sup>4</sup> So if one theory is reducible in the neo-Nagelian sense to another, then the first theory can be explained by the other theory – that is, the behavior of the entities described by the first theory can be explained entirely by appeal to the behavior of the entities described by the second theory. The neo-Nagelian view of reducibility also articulates a second intuition about reducibility, not to be confused with the first intuition, even though it also concerns explanation. The second intuition is that, if a theory is reducible, then, at least in principle, it can be relieved of all its explanatory duties – and thus rendered explanatorily dispensable. The neo-Nagelian view articulates this intuition because, if one theory is reducible in the neo-Nagelian sense to another theory, then any event that could previously have been explained by subsumption under a law of the first theory must also be subsumable under, and hence explainable by, whatever law of the second theory it is from which (given the pertinent type-identity statements) the law of the first theory is deducible; in which case all the explanatory work that was previously done by the first (reducible) theory can in principle be done by the second (potentially reducing) theory.

Type-physicalism was abandoned because it was thought unable to overcome two main objections.<sup>5</sup> Of these, the more influential was that the mental-to-neurophysiological type-identity claims to which it was committed had empirical consequences that were very likely to be false. Suppose, for example, that pain literally is a certain type of (human) neurophysiological state. Then every time anything is in pain, it must be in that very type of neurophysiological state. But both the plasticity of the human brain (e.g., in recruiting new regions of the brain to subserve functions previously subserved by an injured region) and the existence of

pain-feeling creatures with nonhuman neurophysiologies (e.g., octopuses) make it unlikely that any *single* type of human neurophysiological state is required for pain (given that the disjunction of two neurophysiological types doesn't itself count as a single disjunctive neurophysiological type). Contrary to type-physicalism, then, it seems plausible that mental states are 'multiply realized' by many different types of physical state. This, of course, is the famous *multiple realization* objection (see Block and Fodor 237–40; Putnam, 'Nature of Mental States').<sup>6</sup> The second main objection to type-physicalism we may call the *explanatory autonomy* objection. According to it, if, as type-physicalism implies, everyday discourse about mental states is reducible in the neo-Nagelian sense to neurophysiology, then everyday discourse about mental states no longer enjoys explanatory autonomy – it throws no distinctive explanatory light of its own. But everyday mentalistic discourse enables us to express explanatory generalizations that we couldn't express in the proprietary language of neurophysiology. So type-physicalism can't be true (see Putnam, 'Philosophy and Our Mental Life'; Fodor 123–7, 130–1). A third objection to type-physicalism, more influential in the UK than in the US, was Donald Davidson's argument for the thesis that mental phenomena (or at least mental phenomena involving propositional attitudes) are *anomalous*, i.e., that 'there are no strict deterministic laws on the basis of which mental events can be predicted and explained', hence that there are no 'strict laws connecting the mental and the physical', and hence that no 'mental phenomena can be given purely physical explanations' (208, 212, 214). But I will ignore this argument here, since, given its obscurity, adequate discussion would require an article to itself.<sup>7</sup>

### 3. *Event Identity, Supervenience, Realization*

Attempted formulations of non-reductive physicalism have been variations on one of three themes. The first is the *event-identity* theme. This theme appears in Donald Davidson's anomalous monism, which claims (i) that every individual mental event is identical with some or other individual physical event, but (ii) that mental phenomena are anomalous (Davidson 209–15). If the anomalism of mental phenomena rules out reductionism, then claim (ii) makes anomalous monism strongly non-reductive, at least as regards propositional attitudes. The event-identity theme also appears in Jerry Fodor's account of the relationship between physics and the special sciences. This account holds (i) that, as a general rule, no special scientific kind is a physical kind, so that (given the neo-Nagelian view of reducibility) the special sciences aren't in general reducible to physics, but (ii) that every individual event (including therefore every individual mental or psychological event) is identical with some or other individual physical event, and (iii) that all the mechanisms whereby law-like regularities among special scientific phenomena hold are physical (Fodor).

Despite the extensive discussion they've generated, neither anomalous monism nor Fodor's account of the special sciences in relation to physics has much promise as a formulation of non-reductive physicalism. The mental-to-physical event-identity claims that these two views are committed to give rise to a dilemma: either these claims aren't sufficient for physicalism or else they're unintelligible. The first horn is an argument of Kim's. A mental event, for Davidson, is just an event that meets a mental description (i.e., one assumes, has a mental property), and a physical event is just an event that meets a physical description (i.e., one assumes, has a physical property). On this Davidsonian view of events, then, it's easy to see how mental-to-physical event-identity claims could be true even though the mental is irreducible to the physical: if a given mental event is identical with a given physical event, we simply have a single event with both a mental property and a physical property. The rub, however, is that the event-identity claim tells us nothing about the relations between the single event's having the mental property that it has and its having the physical property that it has; a fortiori, the event-identity claim yields no sense in which the single event's having the mental property that it has is *nothing over and above* its having the physical property that it has (see Kim, *Mind in a Physical World* 4–5; see also Horgan, 'Supervenience and Microphysics' 31).<sup>8</sup> Here is the second horn. Suppose, in place of the Davidsonian view of events, that mental events are instantiations of mental properties, and that physical events are instantiations of physical properties (see Kim, *Mind in a Physical World* 121–2, note 7). In that case, the claim that a given mental event is identical with a given physical event must be the claim that a certain instantiation of mental property M just is a certain instantiation of physical property P. But I strongly doubt that sense can be made of such a claim, *given* the need, in the context of trying to formulate non-reductive physicalism, to avoid the physical reducibility of the mental. Such a claim makes perfect sense, of course, if M is the very same property as P; but avoiding the physical reducibility of M requires at least that M not be the same property as P (even if P is just a disjunction of physical properties). So how else to make sense of the claim? The only possibility I know is to treat M as a *functional* property that is *realized* by distinct physical properties, including P, on various occasions. But to follow this approach is to move beyond the event-identity theme to the *third* theme that I'll discuss in attempted formulations of non-reductive physicalism – the realization theme.

Of course, even if the event-identity claims in question, where intelligible, aren't sufficient for physicalism, it doesn't yet follow that either Fodor's or Davidson's *total* view isn't sufficient for physicalism. But although Fodor's total view includes the claim that all the mechanisms whereby law-like regularities among special scientific phenomena hold are physical, his elaboration of this claim commits him to nothing stronger than 'token event identities', which we've already seen are either insufficient for

physicalism or unintelligible (129). And although Davidson proposes adding to anomalous monism the claim that ‘mental characteristics are in some sense dependent, or supervenient, on physical characteristics’ (214), the supervenience of the mental on the physical is the topic of the next several paragraphs, and we will see there what difference, if any, this addition would make.

The second theme discernible in attempted formulations of non-reductive physicalism is the *supervenience/determination* theme, introduced into the philosophy of mind by Davidson, but explored most fully by Jaegwon Kim (*Supervenience and Mind*). Indeed, the vast majority of attempts to formulate non-reductive physicalism have been variations on this theme (see Horgan, ‘Supervenience and Microphysics’, Haugeland, Post, Kirk, and Chalmers 32–42; Jackson 6–27). The intuition behind these attempts is that, if physicalism is true, then the facts about what physical entities there are, what physical properties the physical entities have, what physical relations hold among them, and what physical laws govern them *fix* or *determine* the mental facts – what mental entities there are, what mental properties they have, and so forth. Technical claims of supervenience are used to make precise the idea that the physical facts fix or determine the mental facts, and are then presented as formulations of non-reductive physicalism. To accommodate mental states whose content is wholly or partly determined by conditions external to the thinker, these formulations are typically claims of so-called *global* supervenience, i.e., claims to the effect, roughly, that any two possible worlds (within a certain class) that are exactly alike physically are exactly alike mentally. In all formulations, the class of possible worlds quantified over contains more than the nomologically possible worlds, but, in nearly all formulations, the class contains fewer than the metaphysically possible worlds. Fortunately, however, we need not master the niceties of the many global supervenience claims that have been proposed as formulations of non-reductive physicalism (for a taste, see Horgan, ‘Supervenient Qualia’ 491).

Claims of global supervenience – perhaps unlike other claims of supervenience – appear to be non-reductive because they don’t entail physical-to-mental bridge laws.<sup>9</sup> But precisely because they only assert a very loose connection between the physical and the mental, they confront an objection. Jaegwon Kim has charged claims of global supervenience with being too permissive: the claim that possible worlds exactly alike physically are exactly alike mentally permits a possible world that differs physically from the actual world only in an utterly trivial way (e.g., it has one extra hydrogen atom), but that contains no conscious creatures at all (Kim, ‘“Strong” and “Global” Supervenience Revisited’ 85–6; for thorough discussion, see Post ““Global” Supervenient Determination”).

To my mind, however, the gravest problem with supervenience-themed attempts to formulate non-reductive physicalism is that, although claims of supervenience may be necessary for physicalism, they aren’t sufficient

for it. Recognition of this problem is implicit in Terry Horgan's demand that formulating physicalism requires not just supervenience but superdupervenience, i.e., 'ontological supervenience that is robustly explainable in a materialistically acceptable way' (Horgan, 'From Supervenience to Superdupervenience' 577); he notes, for instance, that G. E. Moore took moral goodness to be non-natural *and* to supervene on natural properties (560–1). Recognition of the problem is also implicit in an important objection that Kim has made over the past decade and a half to supervenience-themed attempts to formulate non-reductive physicalism. Kim's objection is that mental properties, being linked only by supervenience to the physical properties that underlie them, are too loosely connected to physical properties for instances of mental properties to count as causes, given the plausible closure principle that every instance of a physical property has a sufficient physical explanation (see Kim, 'Non-Reductivist's Troubles' and, most recently, *Physicalism, or Something near Enough* 32–69). Now, the goal of Kim's objection has been to cast doubt on the *truth* of the formulations he targets, not on their adequacy as articulations of physicalism. However, he does in fact cast doubt on their adequacy as articulations of physicalism. For he sees himself as urging against supervenience-themed attempts to formulate non-reductive physicalism essentially the same objection that he urges against the explicitly property-dualist view known as *emergentism* – the view that all mental entities are exhaustively decomposable into physical parts, but that at least some of the mental properties of those entities are non-physical and emergent, i.e., they come to be possessed by those entities in accordance with fundamental laws of emergence to the effect that whenever an entity acquires such-and-such a physical character, it thereby comes to possess so-and-so mental property. And the reason why supervenience-themed attempts to formulate non-reductive physicalism are open to this same objection is that they don't bind mental properties to the physical properties that underlie them any more intimately than does emergentism, which actually entails the falsity of physicalism.

That supervenience-themed attempts to formulate non-reductive physicalism aren't sufficient for physicalism can also be argued for directly. The claim that the mental supervenes on the physical is a modal claim and nothing more; it says, roughly, that the way things are physically *necessitates* the way they are mentally. But it doesn't say what explains this necessitation or even that it has an explanation. (The identity of every mental type with a physical type would explain the necessitation; but it would also be reductive.) So the claim that the mental supervenes on the physical is consistent with the possibility that the necessitation of the mental by the physical is just the holding of a brute modal relation between one kind of phenomenon and another, entirely distinct kind of phenomenon. But if the necessitation of the mental by the physical were just the holding of a brute modal relation between physical phenomena and mental phenomena, then there would be no sense in which the

mental was nothing over and above the physical, no sense in which true descriptions of mental affairs were made true by the antics of physical things, and no sense in which the mental was wholly constituted by the physical. So the claim that the mental supervenes on the physical is consistent with the falsity of physicalism as conceived intuitively, and therefore fails to suffice for its truth. It might be objected that the very idea of brute necessitation invoked in the argument is incoherent, and hence that, if the mental supervenes on the physical, there must be an explanation for the necessitation of the mental by the physical. Perhaps so. But all the explanations of the necessitation of the mental by the physical that have ever been suggested are – arguably – reductive (not in the neo-Nagelian sense, but in the sense that the neo-Nagelian sense aims to articulate; see below). So, on pain of lapsing into reductive physicalism, anyone attempting a supervenience-themed formulation of non-reductive physicalism must show how this necessitation can be explained in a non-reductive way (see Melnyk, *Physicalist Manifesto* 58–69). A similar argument against a supervenience-themed formulation of non-reductive physicalism can be made by appeal to non-Humean necessitarianism about laws of nature (see Wilson).

The third theme discernible in attempted formulations of non-reductive physicalism is the *realization* theme, introduced by Richard Boyd (but see also Lycan). He argued that types of mental state aren't identical with types of physical state (because types of mental state are multiply realized) and that individual mental events aren't identical with individual physical events (because they don't share their essential properties). But these failures of mental-to-physical identity don't mean that physicalism is false, he said, because it's sufficient for physicalism if 'in the actual world all mental phenomena are *physically realized*' (Boyd 87; my italics).

Boyd's appeal to physical realization is intriguing, but hard to evaluate, because he offered no details. For example, he gave no account of realization, no doubt assuming his readers would understand what he meant from such paradigm cases of realization as the relationship between corkscrews and the different physical structures that constitute them or the relationship between animal legs (e.g., mammal and insect) and the different structures that constitute them. But a formulation of physicalism that appeals to realization can be worked out in detail (see Melnyk, *Physicalist Manifesto* ch. 1). However, the result – realization physicalism – is, I believe, a form of physicalism that, contrary to Boyd's hope and expectation, is *reductive*. Recall that the neo-Nagelian view of reducibility articulates the intuition that reduction is the synchronic, non-causal explanation of the behavior of entities in one theoretical domain entirely in terms of the behavior of entities in another theoretical domain. Now I don't claim that realization physicalism is reductive in the sense of entailing the reducibility of the mental to the physical in the *neo-Nagelian* sense.<sup>10</sup> But I do claim that realization physicalism is reductive in the sense of the intuition, i.e.,



reductive in the sense that, if it's true, then mental phenomena can in principle be synchronically, non-causally explained entirely in terms of physical phenomena. (By 'synchronically explained' I just mean that at least some of the explaining physical phenomena are simultaneous with the mental phenomena being explained.) An obvious corollary of my claim is that the neo-Nagelian view of reducibility isn't the only possible articulation of the core intuition that reduction is the synchronic, non-causal explanation of phenomena of one kind entirely in terms of phenomena of another kind.

To see why realization physicalism is reductive, we need first to understand more clearly what it says. Roughly, it says that every mental state-token is either (1) a token of a physical state-type or (2) a *physically realized* state-token of some *functional* state-type. Condition (2) obviously needs explanation. A 'functional' state-type is here construed very broadly; it's meant to be a kind of state-type such that something's being in that state-type just is its being in some state-type or other that meets a certain specification. The composition and working of the 'some state-type or other' don't matter, so long as it meets the specification in question; and the specification could be of any sort – it might be the playing of a causal role, as traditionally thought, but it needn't be.<sup>11</sup> A car engine's being on is an example of a functional state: for a car engine to be on is just for it to be in some state-type or other that explains why, if the containing car is in gear, its brakes off, and its gas pedal depressed, it then moves. Perhaps having an ache in the shoulder is a functional state – one that someone is in just in case he or she internally represents damage in his or her shoulder region, but does so by using a system of mental representation importantly different from the one used when *thinking* that his or her shoulder region is damaged.<sup>12</sup> A token of a functional state-type is *physically realized* when, and only when, (i) it's realized by a token of a physical state-type, and (ii) this physical state-type meets the specification in question solely in virtue of the holding of physical laws and perhaps of other physical conditions. So my car engine's being on right now, for example, is very plausibly regarded as physically realized: the type of state it's in that explains why, if it's in gear, its brakes off, and its gas pedal depressed, it then moves is an enormously complex *physical* state-type, which meets the specification just given (i.e., it explains why, if it's in gear *etc.*) solely in virtue of physical laws and other physical states.

Why is realization physicalism reductive? Suppose that a token, *m*, of a mental state-type, M, is physically realized. According to the preceding paragraph, it follows that M just is a functional state-type – call it 'F' – that's tokened iff there's a token of some or other state-type that meets a certain specification – call this specification 'S'. It also follows that there's a token, *p*, of some physical state-type, P, that realizes *m* because P meets S – and does so solely in virtue of physical laws and other physical states. But given these implications of *m*'s being physically realized, it must be

possible to explain *m* entirely in terms of physical phenomena, as follows. Since *p*, and other physical states, exist, and the laws of physics hold, *p* must meet *S*. But there being a state-token of some state-type that meets *S* just is there being a token of *F*, and hence of *M* (since  $F = M$ ), and (we may assume) *this* token of *M* just is *m*. So, given the way things are physically, *m* must exist. Thus, on the assumption that *m* is physically realized, *m* can in principle be explained entirely in terms of the way things are physically. And, by an extension of the same reasoning, if all mental states are physically realized, then – arguably – the holding of regularities among them can also be explained entirely in terms of the way things are physically (for the full story, see Melnyk, *Physicalist Manifesto* 88–110).<sup>13</sup>

But if realization physicalism is reductive, then how does it avoid the two main objections that did for (reductive) type-physicalism? According to the multiple realization objection, of course, it's unlikely that a single type of physical state is required for each type of mental state, contrary to what's implied by the thesis that every type of mental state is identical with some type of physical state. But realization physicalism doesn't require a single type of physical state for each type of mental state; so it doesn't require what's supposed likely to be false. It does require, of course, that every type of mental state be identical with some type of *functional* state; but this is consistent with multiple realization, because very different types of physical state might meet the specification associated with each type of mental (i.e., functional) state, so that mental states are multiply realized.

According to the explanatory autonomy objection, if everyday discourse about mental states is reducible to physics, then no room is left for such discourse to throw any distinctive explanatory light of its own – though in practice it seems to do so. But this objection conflates two distinct ideas, first, the idea that everyday talk about the mind provides *genuine* explanations and, second, the idea that everyday talk about the mind provides *indispensable* explanations. Realization physicalists can allow that everyday talk about the mind provides *genuine* explanations, by insisting that a single mental or behavioral phenomenon can have *both* a reductive explanation in terms of physical events *and* a non-reductive, causal explanation in terms of mental events – just as a window's shattering can have both a reductive explanation in terms of microphysics and a non-reductive, causal explanation in terms of a stray baseball. What makes this possible is that a non-reductive, causal explanation of some phenomenon and a physically reductive explanation of the same phenomenon don't cite the very same events; rather, they fit the phenomenon to be explained into *different* patterns of (physical or physically realized) events. And since two kinds of explanation can throw genuine explanatory light on the same phenomenon, a physically reductive explanation of something doesn't crowd out a non-reductive, causal (e.g., mental) explanation of the same

thing. Obviously, however, a non-reductive, mentalistic explanation can only ever cite mental events that are physically realized, or else physicalism won't be true. So there must inevitably be some sense in which, if realization physicalism is true, perfectly genuine mentalistic explanations can't be *indispensable*; they must in principle be replaceable with explanations couched in the language of physics. And if they aren't replaceable in practice, as clearly they aren't, then the reason is pragmatic only, not metaphysical (for amplification of this paragraph, see Melnyk, *Physicalist Manifesto* 164–74).

My argument that realization physicalism is reductive has, of course, depended on construing reducibility in a certain way. Now, I don't expect anyone to deny that this is one legitimate way of construing reducibility. But some philosophers insist that there is at least one other kind of reducibility. For example, Kim remarks: 'physically irreducible properties remain outside the physical domain – that is, if anything is physically reduced, it must be identical with some physical item' (*Physicalism, or Something near Enough* 34). What then? If Kim is right that there's a kind of physical reduction that requires identity with some physical item (a physical token would presumably do), then realization physicalism must be rated reductive in one important way, but weakly non-reductive in another. For the claim that a mental state-token is physically realized by a physical state-token apparently doesn't entail that the mental state-token is identical with the physical state-token. And, indeed, if Boyd was right to argue that a mental state-token and its physical realizer don't share all their modal properties, then they aren't identical. However, I doubt that Kim is right. The kind of physical reducibility that Kim is talking about, following Smart, is the property of being nothing over and above the physical. But if realization physicalism is correct, then, *pace* Kim, mental items needn't be identical with physical items in order to be nothing over and above the physical; they can be distinct from all physical items, but still nothing over and above the physical, just so long as they are physically realized.

#### 4. Pereboom on the Formulation of Non-Reductive Physicalism

Derk Pereboom has recently tried to formulate a version of physicalism that's non-reductive – non-reductive in the strong sense of entailing the irreducibility of mental phenomena to physical phenomena. After explaining why Pereboom thinks his formulation is physicalist and yet still strongly non-reductive, I'll explain why I doubt that he's right.

Pereboom explicitly denies that mental event-types are one and the same as microphysical event-types (500), and that mental event-tokens are one and the same as microphysical event-tokens (503).<sup>14</sup> To articulate the idea that mental phenomena are nothing over and above physical phenomena, he appeals instead to *constitution*, claiming that every mental event is

'wholly constituted of' some or other microphysical event (500, 526–7). Unfortunately, however, Pereboom offers no account of constitution. All he says is that, if a physical event-token constitutes a mental event-token, then the physical event-token, 'together with any requisite relational features', will be 'sufficient' for the mental event-token (500). But, for all that Pereboom says, this sufficiency might be sufficiency in accordance with a fundamental law of physical-to-mental emergence whereby, if an event of *p*'s physical type occurs, then an event of *m*'s mental type occurs; and if it is, then (mental) *m*'s being constituted by (physical) *p* won't entail that *m* is nothing over and above the physical. So Pereboom needs to say more about constitution.

Perhaps he could do so by simply identifying *constitution* with the relation of *realization* as this is understood by realization physicalists. For suppose that event *m* is of mental event-type M, and that it's constituted – in the realization physicalists' sense of realized – by physical event *p*. Then, as we saw above, given the occurrence of *p*, the physical laws, and perhaps other physical events, *m* must occur. But the reason *m* must occur, given these physical conditions, is that (i) *p* is an event of physical event-type P, (ii) physical event-type P meets so-and-so specification solely in virtue of physical laws and perhaps other physical events, and (iii) the occurrence of an event of type M just *is* the occurrence of an event of some type or other that meets so-and-so specification.<sup>15</sup> Once this explanation is appreciated, it's clear that *m* is nothing over and above the physical.

At first sight, however, it appears that Pereboom is debarred from identifying constitution with realization, because he explicitly rejects a functionalist account of mental events (511ff.), and only functional events are candidates for realization in the realization physicalists' sense. But closer examination reveals that Pereboom only means to reject one form of functionalism, the form according to which the occurrence of an event of mental type M is the occurrence of an event of some type or other that is related thus and so (e.g., causally) to other events of other types (see, e.g., 515); we might call this 'external-relations functionalism'. More importantly, Pereboom's proposed alternative says that the occurrence of an event of mental type M is the occurrence of an event of some type or other that has a certain *internal structure* (517–19). He explains it by analogy with a ball-piston engine, saying that 'Characteristic of this engine is its having parts with particular shapes and rigidities, and these parts must be arranged in a particular way' (517); and he stresses that a ball-piston engine might be physically realized by any physical system that has parts with the requisite shapes and rigidities, arranged in the right way, no matter what its fine-grained microphysical character. But Pereboom's proposed alternative is still functionalist in the realization physicalist's sense, since it says that what makes an event into a mental one is its meeting a certain specification – that of having such-and-such an internal

structure. So Pereboom can after all express his physicalism by claiming that all mental events are physically realized.<sup>16</sup>

If he does so, however, then Pereboom's physicalism turns out to entail realization physicalism, and therefore to be reductive in the same sense as realization physicalism is (i.e., if it's true, then mental phenomena can in principle be synchronically and non-causally explained entirely in terms of physical phenomena). And indeed he acknowledges this, I think, when he writes that, according to his formulation, and contrary to emergentism, 'higher-level [e.g., mental] properties are, in fact, predictable from basal [i.e., physical] conditions', i.e., predictable without appeal to fundamental laws of physical-to-mental emergence (507–8).<sup>17</sup> So why does Pereboom deem his formulation of physicalism to be non-reductive? He says it's because his formulation claims that 'mental states instantiate irreducibly mental causal powers' (499); and certainly talk of 'irreducibly mental causal powers' is the sort of talk that quickens anti-reductionist pulses. These causal powers are 'irreducibly mental' since, first, the mental causal power-tokens of a mental state-token aren't of the same types as the physical causal power-tokens of any physical state-token and, second, mental causal power-tokens aren't one and the same as the physical causal power-tokens of any physical state-token (500). If, contrary to this second claim, mental causal power-tokens were identical with certain physical causal power-tokens, then, he says, 'Higher level kinds and explanations would at best group token microphysical causal powers in a way that does not correspond to the classifications of microphysics itself' (502; see also 500); and it's clear that, if that's all they did, then they wouldn't exhibit the sort of irreducibility that Pereboom is looking for.

Now for my doubts about Pereboom's proposed formulation of non-reductive physicalism. It allegedly counts as non-reductive because it claims that 'mental states instantiate irreducibly mental causal powers'. But how in that case can it still count as a version of physicalism? Specifically, what is the sense in which 'irreducibly mental causal powers' are nothing over and above the physical? Pereboom answers that 'token mental causal powers are wholly constituted by token microphysical causal powers' (504); that is, he appeals to his relation of constitution. But, as we've seen, Pereboom's constitution relation must be construed as realization in the realization physicalist's sense, in which case token mental causal powers can be 'wholly constituted by token microphysical causal powers' only if token mental causal powers turn out to be objects of some broadly functional type – or, in case causal powers aren't genuine objects, only if the *property* of having such-and-such a mental causal power turns out to be a broadly functional property. But because Pereboom, like others who appeal to causal powers, says nothing about their nature, one can't even begin to assess the idea that they are broadly functional objects, or that the property of having such-and-such a causal power is a broadly functional property. In consequence, he leaves it open whether his

proposed formulation of non-reductive physicalism succeeds where its predecessors have failed.

### *Acknowledgment*

For very helpful comments on earlier drafts of this paper I am grateful to Brie Gertler and an anonymous referee.

### *Short Biography*

Andrew Melnyk is interested in many aspects of the philosophy of science, and in all aspects of the philosophy of mind, but has published mostly in the metaphysics of mind, his work appearing in such journals as the *Journal of Philosophy*, *Noûs*, *Philosophy and Phenomenological Research*, *Philosophical Studies*, and the *Australasian Journal of Philosophy*. His book, *A Physicalist Manifesto: Thoroughly Modern Materialism*, appeared in 2003 with Cambridge UP. He is presently a professor of philosophy and chair of the department at the University of Missouri. Born in London, Andrew Melnyk was educated at St. Paul's School, London, and Oxford University, where he read Greats before taking a B.Phil. and D.Phil. in philosophy.

### *Notes*

\* Correspondence address: Department of Philosophy, 439 Strickland Hall, University of Missouri-Columbia, Columbia, MO 65211. Email: melnyka@missouri.edu.

<sup>1</sup> The term, 'physicalism', is often used to refer to the comprehensive thesis that (very roughly) all phenomena – not just mental phenomena – are nothing over and above physical phenomena.

<sup>2</sup> I have extensively discussed these questions, and others, elsewhere (Melnyk, *Physicalist Manifesto*).

<sup>3</sup> The modification was to insist, as Nagel himself did not, that bridge laws take the form of type-identity claims, as opposed, for example, to claims of nomological co-extensiveness.

<sup>4</sup> That is, some variation on the theme of the deductive-nomological account of explanation seems plausible in the case of laws (and hence of theories).

<sup>5</sup> An objection I omit from the text is that, because the representational content of a mental state is determined, at least in part, by non-neurophysiological circumstances *external* to the head of the mental state's bearer, mental states with representational content can't be identical with types of neurophysiological state (see Putnam, 'Meaning of "Meaning"'). I omit this objection because, although it rules out the type-identity of mental states with neurophysiological states, it doesn't rule out the type-identity of mental states with neurophysiological-cum-environmental states, and so is consistent with reductive type-physicalism (unlike the objections in the text).

<sup>6</sup> However, not everyone agreed, or agrees, that the mental is multiply realized by the physical (see, e.g., Shapiro). Also, though usually called the 'multiple realizability objection', it only appeals to (alleged) cases of *actual* multiple realization. A genuine multiple realizability objection would appeal to the metaphysical *possibility* of, say, pain without the neurophysiological state thought to be identical with pain; such a bare possibility would be enough to contradict the proposed identity claim, given the necessity of identity.

<sup>7</sup> Jaegwon Kim has made an interesting and very sympathetic attempt to understand it (Kim, 'Psychophysical Laws').

<sup>8</sup> Elsewhere I've argued that the event-identity claim doesn't entail the supervenience of the mental on the physical, something that any version of physicalism should entail (Melnyk,

*Physicalist Manifesto* 68–9). Interestingly, Kim has made an apparently opposite objection to anomalous monism, namely, that it has the presumably unintended effect of denying the reality of the mental (Kim, 'Myth of Non-reductive Materialism' 267–71).

<sup>9</sup> Kim has argued that claims of so-called *strong* supervenience turn out to entail the existence of law-like connections between mental and physical properties that might enable the reduction of the mental to the physical (Kim, 'Supervenience as a Philosophical Concept' 151–4).

<sup>10</sup> So if this is the only kind of reducibility Boyd had in mind, which is possible, he might agree with me.

<sup>11</sup> Nothing in the realization physicalist's account of realization – and nothing in realization physicalism more generally – requires that the functional nature of any non-physical type be discoverable a priori. In particular, realization physicalism isn't a semantic thesis; it neither asserts nor requires the existence of any functional *concepts* or *predicates*. So if any non-physical types (e.g., folk psychological, biological, or geological types) turn out to be identical with functional types (in the very liberal sense intended), then those identities will have to be discovered a posteriori.

<sup>12</sup> The representational theory of phenomenally conscious mental states mooted here is developed and defended by Michael Tye.

<sup>13</sup> The style of reductive explanation here described was first noted at about the same time by several philosophers (see Brooks; Chalmers 42–7; Kim, *Mind in a Physical World* 97–103; Melnyk, 'Two Cheers For Reductionism').

<sup>14</sup> He argues for the distinctness of mental from physical event-tokens by noting their differing essential properties.

<sup>15</sup> I have simplified the reasoning a little to ease the exposition.

<sup>16</sup> Indeed, he sometimes uses the language of realization (e.g., 504).

<sup>17</sup> In the same passage, he makes an exception for higher-level properties that are relational; but presumably these too will be predictable from basal conditions if basal conditions are expanded beyond basal constituters to include basal environmental conditions.

### Works Cited

- Block, Ned and Jerry Fodor. 'What Mental States Are Not'. *Readings in the Philosophy of Psychology*. Vol. 1. Ed. Ned Block. London: Methuen, 1980. 237–50. [Originally: *Philosophical Review* 81 (1972): 159–81.]
- Boyd, Richard. 'Materialism Without Reductionism: What Physicalism Does not Entail'. *Readings in the Philosophy of Psychology*. Vol. 1. Ed. Ned Block. London: Methuen, 1980. 67–106.
- Brooks, David. 'How to Perform a Reduction'. *Philosophy and Phenomenological Research* 54 (1994): 803–14.
- Chalmers, David. *The Conscious Mind: In Search of a Fundamental Theory*. New York, NY: Oxford UP, 1996.
- Davidson, Donald. 'Mental Events'. *Essays on Actions and Events*. New York, NY: Oxford UP, 1980. 207–25.
- Fodor, Jerry. 'Special Sciences, or the Disunity of Science as a Working Hypothesis'. *Readings in the Philosophy of Psychology*. Vol. 1. Ed. Ned Block. London: Methuen, 1980. 120–33. [Originally: *Synthese* 28 (1974): 97–115.]
- Haugeland, John. 'Weak Supervenience'. *American Philosophical Quarterly* 19 (1982): 93–103.
- Hellman, Geoffrey and Frank Thompson. 'Physicalism: Ontology, Determination, and Reduction'. *Journal of Philosophy* 72 (1975): 551–64.
- Horgan, Terence. 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World'. *Mind* 102 (1993): 555–85.
- . 'Supervenience and Microphysics'. *Pacific Philosophical Quarterly* 63 (1982): 29–43.
- . 'Supervenient Qualia'. *Philosophical Review* 96 (1987): 491–520.
- Jackson, Frank. *From Metaphysics To Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford UP, 1998.
- Kim, Jaegwon. *Physicalism, or Something near Enough*. Princeton, NJ: Princeton UP, 2005.
- . *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press, 1998.

- . 'The Myth of Non-Reductive Materialism'. *Supervenience and Mind: Selected Philosophical Essays*. New York, NY: Cambridge UP, 1993. 265–84.
- . 'The Non-Reductivist's Troubles with Mental Causation'. *Supervenience and Mind: Selected Philosophical Essays*. New York, NY: Cambridge UP, 1993. 336–57.
- . 'Psychophysical Laws'. *Supervenience and Mind: Selected Philosophical Essays*. New York, NY: Cambridge UP, 1993. 194–215.
- . "'Strong" and "Global" Supervenience Revisited'. *Supervenience and Mind: Selected Philosophical Essays*. New York, NY: Cambridge UP, 1993. 79–91.
- . *Supervenience and Mind: Selected Philosophical Essays*. New York, NY: Cambridge UP, 1993.
- . 'Supervenience as a Philosophical Concept'. *Supervenience and Mind: Selected Philosophical Essays*. New York, NY: Cambridge UP, 1993. 131–60.
- Kirk, Robert. 'How Physicalists Can Avoid Reductionism'. *Synthese* 108 (1996): 157–70.
- Lycan, William. 'Form, Function, and Feel'. *Journal of Philosophy* 78 (1981): 24–49.
- Melnyk, Andrew. *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge UP, 2003.
- . 'Two Cheers for Reductionism; or, the Dim Prospects for Non-Reductive Materialism'. *Philosophy of Science* 62 (1995): 370–88.
- Nagel, Ernest. *The Structure of Science: Problems in the Logic of Scientific Explanation*. Indianapolis, IN: Hackett, 1979.
- Pereboom, Derk. 'Robust Nonreductive Materialism'. *Journal of Philosophy* 94 (2002): 499–531.
- Post, John. *The Faces of Existence: An Essay in Non-reductive Metaphysics*. Ithaca, NY: Cornell UP, 1987.
- . "'Global" Supervenient Determination: Too Permissive?' *Supervenience: New Essays*. Eds. Elias E. Savellos and Ümit D. Yalçın. Cambridge: Cambridge UP, 1995. 73–100.
- Putnam, Hilary. 'The Meaning of "Meaning"'. *Mind, Language, and Reality: Philosophical Papers*. Vol. 2. Cambridge: Cambridge UP, 1975. 215–71.
- . 'The Nature of Mental States'. *Mind, Language, and Reality: Philosophical Papers*. Vol. 2. Cambridge: Cambridge UP, 1975. 429–40.
- . 'Philosophy and Our Mental Life'. *Mind, Language, and Reality: Philosophical Papers*. Vol. 2. Cambridge: Cambridge UP, 1975. 291–303.
- Shapiro, Lawrence. *The Mind Incarnate*. Cambridge, MA: MIT Press, 2004.
- Sklar, Lawrence. 'Types of Inter-Theoretic Reduction'. *British Journal for the Philosophy of Science* 18 (1967): 109–24.
- Smart, J. J. C. 'Sensations and Brain Processes'. *Philosophical Review* 68 (1959): 141–56.
- Tye, Michael. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press, 1995.
- Wilson, Jessica. 'Supervenience-Based Formulations of Physicalism'. *Noûs* 39 (2005): 426–59.