

*If You Can't Reduce,
You Must Eliminate*

Why Kim's Version of Physicalism
Isn't Close Enough

A Review Essay on
Physicalism, or Something Near Enough

J. P. MORELAND
*Talbot Department of Philosophy
Biola University
La Mirada, California*

Physicalism, or Something Near Enough. By Jaegwon Kim. Princeton: Princeton University Press, 2005. 186 pages. \$26.95.

Kim's book reads like a battle-tested, veteran general, providing a summary to the combatants of sixty years of warfare, along with an assessment of the current battlefield and a declaration of the only way forward. His central thesis—pronouncement might be better—is this: Given that physicalism is the only credible worldview we have, the only way to preserve mental causation is to reduce functionally the mental to the physical. Substance dualism cannot account for mental causation, Nagelian reduction is dead, and type physicalism is just false when it comes to qualia. Cognitive mental states can be functionally reduced, so cognition and agency can be saved, but the intrinsic features of qualia cannot. We can functionally reduce the relational aspects of qualia used to sort and discriminate among kinds of qualia, but their intrinsic features are irreducible. Since qualia are real yet epiphenomenal, we are stuck with them, and that means the physicalism *tout court* fails. But the failure is one we can live with and we end up with a form of physicalism that is close enough.

The book is an updated version of his 1998 work with MIT Press titled *Mind in a Physical World*. Heretofore, I have used the earlier book as a text,

and it would be a mistake to think that this new work renders *Mind in a Physical World* obsolete. In the earlier work, Kim deftly describes the relationship among supervenience, realization, and functional reduction, he provides detailed analysis of the mereological hierarchy, he clarifies and allegedly solves the generalization problem (given that mental causation is impossible in light of the hierarchy, it would seem that causal impotence affects all properties unique to any special science). None of these is treated with the same focus in the new book. On the other hand, Kim's new offering provides an updated and more powerful presentation of the case against mental causation, he addresses substance dualism and mental causation, offers updated and powerful critiques of type physicalism, and explicitly addresses the sort of qualia dualism with which a physicalist can live.

In chapter 1, Kim lays out the two central physicalist problems with the mental. He claims that while the identification and resolution of the core issues in philosophy of mind are dependent on one's worldview, given that physicalism is the only credible worldview we have, the two key issues in the mind/body problem are mental causation and the existence of phenomenal consciousness. Kim correctly notes that the problem of mental causation arises from the very nature of physicalism itself, and not from a Cartesian view of mental substance, and, indeed, mental causation is a difficulty in the category of property every bit as much as in the category of substance.

The supervenience argument (also known as the exclusion argument) shows that there can be no mental causation in a world that is fundamentally physical, and this raises serious problems regarding cognition and agency, two features of our lives that are hard to give up. The supervenience argument, says Kim, may be construed to show that mental causation is inconsistent with the conjunction of four theses: (1) closure; (2) exclusion (no overdetermination); (3) supervenience (not construed simply as property covariance, but taken to entail dependence and synchronic determination); (4) mental irreducibility.

Moreover, qualia cannot be reduced to physical phenomena because (a) Nagelian bridge laws are not available (for example, multiple realization issues) and, even if they were, they cannot be turned into identities. (b) Functional reduction is the only plausible way to reduce the mental but qualia are paradigm cases of entities essentially characterized by their intrinsic and not their relational features, so functional reduction for qualia is a nonstarter. So we are stuck with a two world-knots: Mental causation can be solved if and only if the mental is functionally reducible, the mystery of consciousness is solvable only if consciousness is functionally reducible, but phenomenal consciousness is not so reducible. Ergo, we are stuck with the problem of phenomenal consciousness.

Kim is one physicalist who faces head-on the limitations inherent in that standpoint, and this first chapter gets right to the point with clarity and force. However, I want to urge caution about a brief assertion in the opening paragraph of the chapter that is made almost in passing, namely, that the precise nature of and possible solutions for the issues constituting the mind/body problem are dependent on the larger philosophical/scientific framework one brings to the mind/body problem. There is a sense in which this is true, namely, that certain phenomena (for example, the existence of consciousness) become problematic for some worldviews and not others.

But in a more important sense, Kim's assertion is false. The core data that constitute recalcitrant facts regarding human persons arise from prephilosophical self-awareness, and they transcend culture or worldview. Average people claim to be aware of themselves as simple, unified centers of consciousness, different from their bodies (or any chunks of matter) and conscious states, who can survive bodily death, who exercise agent causal power for the sake of irreducibly teleological ends, who can know not just what their phenomenological qualia are by direct first-person introspection, but also what their thoughts, beliefs and desires are, and so on. I am not claiming that average folk would speak in all these ways, but I am claiming that given a set of conceptual distinctions, this is what most people would claim to believe about themselves on the basis of self-awareness.

Of course, they could be wrong. But two things should be kept in mind. First, this evidence of so-called folk ontology is relatively invariant across worldviews. Second, these are features that, as Chisholm put it, we have a right to believe about ourselves. By framing the mind/body problem as largely worldview dependent, Kim, perhaps inadvertently, distorts the nature of the mind/body problem in a way that serves physicalism. More generally, Kim's approach is a substitution instance of the more general way in which philosophical naturalism has distorted the way metaphysics, including philosophy of mind, should be done. I know my claims are controversial, and I cannot say more about them here. But I do not want a seemingly brief methodological and epistemic point at the beginning of such an important book to go unnoticed.

In chapter 2, Kim refines and clarifies his earlier versions of the supervenience argument, and responds to two criticisms of it: Kim's dismissal of overdetermination and the problem of causal drainage. The fundamental idea of the supervenience argument is that "vertical determination excludes horizontal causation." To see this, Kim invites us to consider two physical events, p and p^* , along with two mental events, m and m^* such that (1) m and m^* supervene on p and p^* respectively (where supervenience includes the notion of dependence and determination, even if this is not taken to be efficient causality) and (2) p causes p^* .

The argument proceeds in two stages. Stage 1: Focus on m to m^* causation. Since m^* obtains in virtue of p^* , if m is going to cause m^* it must do so by causing p^* . Stage 2: Kim offers two different ways to complete the argument. Completion 1: Assuming causal closure and exclusion (no causal overdetermination), p will be the cause of p^* and there is no room for m to be involved in bring about p^* . We have m and m^* supervenient on p and p^* , respectively, and p causing p^* , nothing more and nothing less. Completion 2: Granting that m causes m^* by causing p^* , it follows that m causes p^* . By irreducibility, we have $m \neq p$. So m and p cause m^* . By exclusion and closure, m is ruled out and p is selected as the only cause for p^* . Completion 2 avoids reference to supervenience and to the assertion that p has a valid claim to be the cause of p^* . On either way of completing stage two, we have $m \rightarrow m^*$ and $m \rightarrow p^*$ giving way to $p \rightarrow p^*$.

Kim then responds to an argument proffered by Tom Crisp and Ted Warfield to the effect that Kim's earlier case against overdetermination fails. Kim had argued that overdetermination should be rejected because it violates physical causal closure on the grounds that in a near world without p but with m , m would cause p^* leaving p^* without a physical cause. Kim acknowledges that as stated, his argument fails because if p is removed in such a world, m goes with it. Kim's previous argument works only if there is a possible world with m and without p , and given supervenience and a form of modality attached to it, the world in question is impossible in the corresponding modal sense.

However, Kim argues that Crisp and Warfield's dilemma-argument against him fails to justify overdetermination. Consider lemma one: Suppose supervenience holds in a world w with m but without p . There will be some other subvenient base for m , p' , which restores closure, so, *contra* Kim, an alleged violation of closure cannot be used as an argument against overdetermination. However, Kim responds that if we imagine a sort of free-riding causal chain from m to p^* riding on p to p^* in the actual world and on p' to p^* in w , there is no way to conceive of how there could be room in either case for a free-riding chain to enter into the production of p^* . For some, this may be a fair rebuttal to Crisp and Warfield if their argument is an attempt to justify overdetermination. But this first lemma does seem to be effective in shifting Kim's justification for rejecting overdetermination from alleged violations of closure to our difficulty of conceiving room for top-down causation given supervenience.

Now consider lemma two: Assume that supervenience does not hold in w . In this case, m does, indeed, cause p^* and closure is violated. But given supervenience, argue Crisp and Warfield, such a world w is impossible in a relevant modal sense. So why should nonreductionists abandon overdetermination because it implies that closure fails in impossible worlds! They should be indifferent to w in this case and retain overdetermination.

Kim responds that all physicalists should want to avoid worlds like this and not be indifferent to them, not because they are nomologically impossible with respect to physical laws, but because they are impossible with respect to psycho-physical laws. This argument requires that physicalists reject closure simpliciter (every physical event that has a cause has a physical cause) in favor of strong closure (a caused physical event *cannot* have a nonphysical cause; it *must* have a physical cause). Kim acknowledges that this stronger view may well be question-begging against those who believe in top-down mental causation, but he believes that there is no good reason for a self-reflective physicalist to reject strong closure.

The chapter closes with Kim responding to an argument designed to show that if successful, Kim's exclusion argument generalizes such that it renders causally impotent all properties of the special sciences. Space forbids me to discuss this dialectic in detail, but two brief points need to be made. If successful, Kim's supervenience argument does, in fact, render epiphenomenal any genuinely emergent property. The only "properties" that may be safe from causal drainage are structural properties.

Second, Kim claims that macroreduction is quite different from macroelimination in the category of property (he advocates the former and rejects the latter). But Kim's solution to causal drainage preserves macrocauses at the expense of performing a reduction of macroentities in the category of substance, and as far as I can see, Kim's views on these matters are no different than eliminativists like Peter van Inwagen or Trenton Merricks. My observation may be taken as a virtue of Kim's approach or as a *reductio* against him, but one thing seems clear. Kim's microreduction in the category of individual means that the mereological hierarchy is no longer a *hierarchy*. One does not move up anything as one moves to macro-wholes. This is clear if such wholes are eliminated. But even if they are not, they turn out to be spatiotemporally wider relational entities located at the bottom level if such there is. They are not higher up.

Given that mental causation is problematic for physicalism, does this mean that substance dualism is back on the table? Kim turns back this move in chapter 3 by offering a critique of mental-to-physical or mental-to-mental causation in a substance dualist framework. On a Cartesian view (minds are nonspatial entities causally connected to their bodies), the problem, says Kim, is not that mind and body are different simpliciter or that a Cartesian cannot just take causal interaction as basic. The problem has to do with the nonspatiality of the Cartesian ego, and it is surfaced by the Cartesian's inability to solve in a noncircular way the pairing problem. I think it is important to keep in mind the spatial problem because it serves as a defeater for standard Cartesian attempts to solve causal pairing issues (for example, singular haecceitistic causation; precisely because it is *sui generis*, causal

bruteness simply disallows the question surfaced in causal pairing). According to Kim, the only intelligible way to depict how A can be a cause of B is if we can trace a continuous causal pathway from A to B each pair of which stands in some causal pairing relation R, or if R just is the relation between A and B. It is clear that the case of remote causes and effects in scenario one reduces to scenario two, so either way, we must identify R such that it renders causality intelligible and provides a solution to causal pairing.

According to Kim, any plausible candidate for R entails space such that the causally connected items must be situated in a space-like framework. Space is essential for causation—spatial structure gives a determinate spatial ordering to entities necessary for causality—and this solves the pairing problem. It also disallows Cartesian egos from entering into causal relations with either matter or other such egos.

Next, Kim argues against dualist attempts to locate souls in space as extensionless points. Among his reasons for rejecting this move is the claim that it makes no sense to say that a soul is spatially in the body but not at any exact point. Moreover, if locating souls in space is to solve the causal pairing problem, it must be the case that no more than one soul can occupy a single spatial point. Further, if an entire soul can be at a geometric point, asks Kim, how can it have enough structure to account for all the marvelous causal work it's supposed to perform?

Since neither version of substance dualism adequately solves the problem of causal pairing, substance dualism must be rejected and we need to try to solve the problem of mental causation within a physicalist framework.

Regarding Kim's critique of Cartesian dualism, it may well be that on a Humean view of causation, spatiotemporal contiguity is a necessary condition for causal interaction, though I am inclined to say that temporal contiguity alone would fit the bill. But the core notion of a more robust realist view of causality is causal production and this, I believe, entails the actualization of a causal power. I see no reason why this core notion or the causal interaction between cause and effect entails spatiality. Indeed, such an entailment renders unintelligible classic understandings of action at a distance, the quest for a cause of the big bang, the claim that, as Plato put it, one form (redness) "brings others with it" (extension) when it is instantiated, and other cases of (alleged) nonspatial causality, and this is surely too strong.

Moreover, there are various ontological "spaces" analogous to ordinary (?) space to serve as adequate grounds for causal "directionality." I cannot develop this point here, but what I think this dialectic comes down to is the intelligibility of nonspatial uses of terms like "in." Aristotle identified eight ways that one thing could be in another and only one was spatial. It is this broader issue, not the role of ordinary space in causality per se that I think is the real issue here.

Finally, it should be said that typical spatial causal connectedness is far from problematic. To cite one example, the notion of one physical object making contact with another one is fraught with difficulties when one tries to analyze it. Indeed, on one such analysis, such contact entails spatially overlapping boundaries between the two objects, an analysis that makes cloudy Kim's appeal to spatial orientation between cause and effect. My point is not that in the absence of an adequate analysis of spatial contact we are not justified in believing in physical causality. Rather, the absence of a complete analysis of such contact afflicts physical causality, and this fact weakens the force of Kim's claim that in the absence of such an analysis regarding Cartesian souls, we are not justified in believing in their existence or causal power.

Regarding Kim's critique of located Cartesian souls, I find his three main arguments weak. Regarding his claim that it is unintelligible to depict a soul as spatially in the body but not at an exact point, it fails for two reasons. First, on the view under consideration, the phrase "not at an exact point" does not mean "not at any point"; it means "not at one point only". It is hard to see why this is unintelligible, especially since many people throughout history, including no small number of philosophers, have taken the omnipresence of God in precisely this sense (though, of course, others reduce God's omnipresence to factors involving His causal and epistemic directness to all spatial locations). The view may be wrong, but it is hard to see why it is unintelligible given that so many find it quite sensible. In fact, Hud Hudson has recently argued that, under certain conditions, composed physical objects may exactly occupy different spatial locations. Even if Hudson is wrong, his view is quite intelligible.

The response might be that it is hard to see how a soul could pick out the arm and cause it to be raised, given that the soul is fully present everywhere in the body. Besides the considerations I raise above in connection with the classic Cartesian view, given that everyone is indexing these days, I do not know why a dualist could not offer an indexing solution to the problem (the soul does not cause the arm to raise simpliciter, but causes-at-location-L the arm to raise or exercises causality L-ly with respect to the arm, or stands in the "causes-at" relation to the arm at L).

Regarding Kim's claim that a located soul solves the pairing problem only if no more than one soul can occupy a single point, this follows only if the fundamental relationship between soul and body is causal, a claim that Thomistic dualists deny.¹ Regarding the problem of a causally relevant structure for an unextended soul, dualists have offered detailed, rich

¹ See J. P. Moreland, Stan Wallace, "Aquinas vs. Descartes and Locke on the Human Person and End-of-Life Ethics," *International Philosophical Quarterly* 35 (1995): 319–30

accounts of such a structure.² Kim is apparently unfamiliar, or at least does not interact with such accounts (Swinburne's book is not included in the bibliography; a sad omission, indeed).

In chapter 4, Kim discusses troublesome questions surfaced by emergentists in the early twentieth century (why do itches and not pains, or why do any mental states, correlate with a given physical state?) that together form the problem of "the explanatory gap." These problematic questions place pressure on physicalists to reduced or at least offer reductive explanations for mental phenomena. Setting aside anomalous monism, I think Kim correctly points out that there are only three noneliminativist options available to physicalists in solving gap problems: Nagelian reduction, functional-realization reduction, or type identity reduction. Nagelian reduction fails due to problems with multiple realization and the fact that contingent bridge laws are just a label for the problem to be solved and not a real solution. According to Kim, functional-realization reduction works for relational mental states that can be functionalized (beliefs, desires, and so on) but not for qualia since their identity is constituted by their intrinsic, not their relational features. This leaves Kripkean identities, and Kim closes the chapter by asking whether or not we should accept these identities for qualia.

In chapter 5, Kim addresses two arguments for why we should take type-type correlations between qualia and physical states to be identities and not mere correlations. One strategy claims that by taking the correlations to be identities we have the best explanation for why the correlations are true, so on the basis of an inference to the best explanation, we should embrace type identity physicalism. Kim's rejoinder is that, in contrast to genuinely contentful deductive-nomological explanations, employing identities and not correlations is explanatorily vacuous and amounts merely to redescribing by way of empty tautologies the very things to be explained. Such "explanations," says Kim, actually dismiss the need for an explanation for such correlations by asserting that no correlations exist in the first place.

The second strategy offered by Block and Stalnaker is an explicitly dismissive one, and Kim favors this approach to strategy one. But Kim still presses the point that, while the dismissive strategy works if we do, in fact, have the relevant identities, that is still an open question which the dismissive strategy fails to answer adequately in the affirmative. However, says Kim, there is a promising line of argument for such identities, and that is to start with acknowledged identities in other areas (water and H₂O, heat and mean kinetic energy) and assimilate mental/physical identities to those. But, Kim argues, the analogy fails. In the case of heat and kinetic energy, the

² See Richard Swinburne, *The Evolution of the Soul*, rev. ed. (Oxford: Clarendon, 1997), chap. 14.

identity is justified because, given physical theory and the identity, we can explain why heating water causes it to boil. Without the identity, we are stuck with “heating water correlates with a phenomenon which causes a phenomenon with which boiling is correlated,” and we should rightly prefer identity to correlation if that avoids such nomological danglers. But given neurophysiological theory and the principle “neural state N_1 causes neural state N_2 ,” if we identify these, respectively, with pain and distress, we do not really explain why pain causes distress, we merely take “pain causes distress” as a rewrite of the neural state principle.

In contrast to what I anticipate will be the response of fellow physicalists, I think Kim is basically correct in chapters 4 and 5, except for one important thing. I do not accept any of the alleged identities usually proffered as analogies for mind/body identities, and limiting myself to those in the category of property, the issue regarding them comes down to weighing these factors: On the side of identity is (1) simplicity and, as Kim correctly notes, (2) the explanatory adequacy of physical theory. On the side of correlation is (1) a realist view of intentionality according to which intentionality is an irreducible feature of mental states that puts a perceiving, knowing subject in epistemically direct and accurate contact with mind-independent objects, and (2) whether some other form of explanation, for example, theistic, is needed to explain adequately the existence and lawlike (seemingly) contingent correlations that obtain in the world.

I will return to this point below in conjunction with my comments on chapter 6. What should be said here, however, is that I do not believe it is as easy as Kim thinks to adopt different positions regarding the various candidates for identities or correlations. Apart from Kripkean considerations about there being no distinction, say, between what a pain is and how it appears, I think that most considerations that favor identifying heat and mean kinetic energy (simplicity, adequacy of physical theory) favor the same for the mind/body identities. And the same goes for adopting these as correlations. By opting for irreducible qualia, Kim’s worldview embraces a range of data for which a correlation alone and not a natural causal explanation is possible. Once the camel’s nose is under the tent, I do not know why one cannot say the same thing for heat and boiling water.

In chapter 6, Kim says that physicalism is at a crossroads, and here is where things stand: We should reject substance dualism, and acknowledge that if mental properties/events are to have causal power, they must be reduced to the physical. This can be done by way of functional-realization reduction for cognitive mental states, says Kim, and thus, we can save agency and cognition.

Crucial to Kim’s view is his conviction that zombie worlds regarding cognitive states are impossible due to Turing test considerations, for example,

from a third-person perspective, it is simply incoherent to withhold belief, knowledge and action ascriptions to zombie creatures. I think Kim's claim here is highly contentious and smacks of a form of antirealist verificationism. After all, from the first-person perspective, there is an obvious difference between a zombie and a real conscious cognizing subject.

Kim would most likely respond that this point incorrectly, even if implicitly, attributes phenomenal qualities to cognitive states, but such qualities should be limited to vivid qualia such as pains. But this claim, which Kim has made elsewhere, seems to presuppose an empiricist sort of "vivacity test" for what counts as having phenomenal qualities so as to qualify as qualia. If there is not an experienceable phenomenological difference between the thought that George Washington is president and the thought, say, that Ben Franklin invented bifocals, how could a person have any idea what he or she was thinking about or what implications do and do not follow from what he or she is thinking about from an introspective, first-person perspective quite prior to there being any publicly observable bodily output?

Returning to Kim's account of where things stand, he opines that even if we can deal appropriately with cognitive mental states, we are still left with a mental residue—qualia. Happily, we can functionally reduce the relational features of qualia relevant to their similarities and differences, and thus, we can save mental causation when it comes to their functionalizable aspects relevant to classifying and sorting them. But we are simply stuck with the intrinsic, epiphenomenal features of qualia. Since there is no credible general worldview besides physicalism, this is close enough to physicalism to be satisfying.

The issue here is the adequacy of physicalism as a worldview. In *From Metaphysics to Ethics*, Frank Jackson was correct to say that various phenomena are on an explanatory par for the physicalist, and they must be either eliminated or reduced: consciousness, indexical facts, normative properties and, yes, secondary qualities. It may be fine and dandy to admit defeat regarding qualia if, indeed, physicalism is the only credible worldview on the table. But as Jackson correctly notes, consciousness, including qualia, secondary qualities and other phenomena all require parallel treatment, more so than Kim acknowledges. For Jackson, we have to reduce or eliminate them one and all.

But if Kim's strategy regarding qualia is accepted, that provides grounds for claiming, on pain of inconsistency, that a whole host of other phenomena are similarly inadequately assimilated into a physicalist worldview. Embarrassing emergentist questions would start popping up all over the place. Some like Searle try to dismiss the embarrassment of such questions for naturalists on the grounds that such contingent facts may be left as

brute. But aside from the fact that these are no longer brute *physical* facts, contingent facts are poor candidates for bruteness, unless, of course, they result from libertarian actions, a point that brings a smile to natural theologians. That is why some who accept such emergent brute facts/properties claim that they are at least nomologically necessitated by their physical base. But this is a hard sell, especially in light of inverted qualia possibilities, possibilities which, to his credit, Kim accepts. Thus many naturalists sensitive to worldview issues follow Jackson: what you cannot reduce you must eliminate.

The admission of qualia is especially troublesome if there are antecedent grounds for accepting theism, and if qualia (secondary qualities, and so on) provide data for an argument for God's existence, an argument that is part of a broader cumulative case. If I am right about this, then I suspect that many physicalists will have the jitters about accepting qualia. Kim has admitted elsewhere, that if a whole range of phenomena is not among basic physical phenomena and those phenomena defy physical explanation, it would be time to reexamine one's physicalist commitments. It may be time for Kim to do just that. Given theism as a plausible rival to naturalism, Kim's version of physicalism may not be close enough after all.