



The Myth of Nonreductive Materialism

Author(s): Jaegwon Kim

Source: *Proceedings and Addresses of the American Philosophical Association*, Vol. 63, No. 3 (Nov., 1989), pp. 31-47

Published by: [American Philosophical Association](#)

Stable URL: <http://www.jstor.org/stable/3130081>

Accessed: 24/03/2011 08:49

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=amphilosophical>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Philosophical Association is collaborating with JSTOR to digitize, preserve and extend access to *Proceedings and Addresses of the American Philosophical Association*.

<http://www.jstor.org>

THE MYTH OF NONREDUCTIVE MATERIALISM

Jaegwon Kim
Brown University

Presidential Address delivered before the Eighty-seventh Annual Central Division Meeting of the American Philosophical Association in Chicago, Illinois, April 28, 1989.

I

Reductionism of all sorts has been out of favor for many years. Few among us would now seriously entertain the possibility that ethical expressions are definable, or reducible in some broader sense, in terms of “descriptive” or “naturalistic” expressions. I am not sure how many of us can remember, in vivid enough detail, the question that was once vigorously debated as to whether so-called “physical-object statements” are translatable into statements about the phenomenal aspects of perceptual experience, whether these are conceived as “sense data” or as some manner of “being appeared to”. You may recall the idea that concepts of scientific theories must be reduced, via “operational definitions”, to intersubjectively performable procedures whose results can be ascertained through observation. This sounded good—properly tough-minded and hard-nosed—but it didn’t take long for philosophers and scientists to realize that a restrictive constraint of this sort was neither enforceable nor necessary—not necessary to safeguard science from the threat of metaphysics and pseudo-science. These reductionisms are now nothing but museum pieces.

In philosophy of mind, too, we have gone through many reductionisms; some of these, such as logical behaviorism, have been defunct for many years; others, most notably the psychoneural identity theory, have been repeatedly declared dead; and still others, such as versions of functionalism, are still hanging on, though with varying degrees of difficulty. Perhaps as a result of the singular lack of success with which our earlier reductionist efforts have been rewarded, a negative image seems to have emerged for reductionisms in general. Many of us have the feeling that there is something rigid and narrow-minded about reductionist strategies. Reductionisms, we tend to feel, attempt to impose on us a monolithic, strait-jacketed view of the subject matter, the kind of cleansed and tidy picture that appeals to those obsessed with orderliness and discipline. Perhaps this

impression has something to do with the reductionists' ritual incantations of slogans like "parsimony", "simplicity", "economy", and "unity", all of them virtues of a rather puritanical sort. Perhaps, too, reductionisms are out of step with the intellectual style of our times: we strive for patterns of life and thought that are rich in diversity and complexity and tolerant of disagreement and multiplicity. We are apt to think that the real world is a messy place and resists any simplistic drive, especially one carried on from the armchair, toward simplification and unification. In fact, the word "reductionism" seems by now to have acquired a negative, faintly disreputable flavor—at least in philosophy of mind. Being a reductionist is a bit like being a logical positivist or member of the Old Left—an aura of doctrinaire naivete hangs over such a person.

At any rate, reductionism in the mind-body problem has been out of fashion for two decades; it has been about that long since the unexpectedly early demise of the psychoneural identity theory, a doctrine advertised by its proponents as the one that was in tune with a world view adequately informed by the best contemporary science. Surprisingly, the abandonment of psychoneural reductionism has not led to a resurgence of dualism. What is curious, at least in terms of the expectations set by the earlier mind-body debates, is the fact that those who renounced reductionism have stayed with physicalism. The distinctive feature of the mind-body theories that have sprung up in the wake of the identity theory is the belief, or hope, that one can be an honest-to-goodness physicalist without at the same time being a reductionist. In fact, a correct and realistic view of science as it is practiced will show us, the new physicalists assure us, that as an account of the "cross-level" relation between theories, classical reductionism is untenable everywhere, not just about the psychophysical relation. The leading idea in all this has been the thought that we can assuage our physicalist qualms by embracing "ontological physicalism",¹ the claim that all that exists in spacetime is physical, but, at the same time, accept "property dualism", a dualism about psychological and physical attributes, insisting that psychological concepts or properties form an irreducible, autonomous domain. The issue I want to explore here is whether or not a robust physicalist can, consistently and plausibly, swear off reductionism—that is, whether or not a substantial form of physicalism can be combined with the rejection of psychophysical reduction.

To lay my cards on the table, I will argue that a middle-of-the road position of the sort just described is not available. More specifically, I will claim that a physicalist has only two genuine options, eliminativism and reductionism. That is, if you have already made your commitment to a version of physicalism worthy of the name, you must accept the reducibility of the psychological to the physical, or, failing that, you must consider the psychological as falling outside your physicalistically respectable ontology. Of course, you might decide to reconsider your commitment to physicalism; but I will not here consider what dualist alternatives there might be which are still live options for us. So if I am right, the choices we face concerning the mind-body problem are rather stark: there are three—dualism, reductionism, and eliminativism.

1. Throughout I will be using "physicalism" and "materialism" (and their cognates) interchangeably; similarly, "mental" and "psychological".

II

Pressures from two sources have been largely responsible, I believe, for the decline of reductionism in philosophy of mind, a decline that began in the late 1960's. One was Donald Davidson's "anomalism of the mental", the doctrine that there are no precise or strict laws about mental events.² According to Davidson, the mental is anomalous not only in that there are no laws relating mental events to other mental events but none relating them to physical events either. This meant that no nomological linkages between the mental and the physical were available to enable the reduction of the former to the latter. The second antireductionist pressure came from a line of argument based on the phenomenon of "multiple realizability" of mental states which Hilary Putnam forcefully brought to philosophical attention, claiming that it directly refuted the reductive materialism of Smart and Feigl.³ Jerry Fodor and others have developed this idea as a general antireductionist argument, advancing the claim that the "special sciences", such as psychology, sociology, and economics, are in general irreducible to physical theory, and that reductive materialism, or "type identity theory", is generally false as a theory about science.⁴ Earlier physicalists would have regarded the irreducibility as evidence showing the mental to be beyond the pale of a scientifically respectable ontology; that is, they would have inferred eliminativism from the irreducibility. This in fact was Quine's response to the problem of intentionality.⁵ But not for the latter-day physicalists: for them, the irreducibility only meant that psychology, and other special sciences, are "autonomous", and that a physicalist can, in consistency and good conscience, accept the existence of these isolated autonomous domains within science.

Let us begin with Davidson. As noted, the anomalism of the mental can be thought of as the conjunction of two claims: first, the claim that there are no purely psychological laws, that is, laws connecting psychological events with other psychological events, and second, the claim that there are no laws connecting psychological events; with physical events. The second claim, which we might call "psychophysical anomalism", is what

2. See Davidson, "Mental Events" in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980). This paper was first published in 1970.

3. See Putnam, "The Nature of Mental States" in *Mind, Language, and Reality: Philosophical Papers*, vol. II (Cambridge: Cambridge University Press, 1975). This article was first published in 1967.

4. Jerry Fodor, "Special Sciences, or the Disunity of Science as a Working Hypothesis", *Synthese* 28 (1974): 97-115. See also Richard Boyd, "Materialism without Reductionism: What Physicalism Does Not Entail", *Readings in Philosophy of Psychology*, ed. Ned Block (Cambridge: Harvard University Press, 1980).

5. As it is the response of some recent eliminativists; see, e.g., Paul Churchland, "Eliminative Materialism and the Propositional Attitudes", *Journal of Philosophy* 78 (1981): 67-90.

underlies Davidson's argument against reductionism. The argument is simple and direct: the demise of analytical behaviorism scotched the idea that the mental could be definitionally reduced to the physical. Further, psychophysical anomalism shows that a nomological reduction of the mental isn't in the offing either. The implicit assumption about reduction in this argument is one that is widely shared: reduction of one theory to another requires the derivation of the laws of the reduced theory from those of the reducer, and for this to be possible, terms of the first theory must be appropriately connected via "bridge principles", with those of the second. And the bridge principles must be either conceptually underwritten as definitions, or else express empirical lawlike correlations ("bridge laws" or "theoretical identities").⁶

This is all pretty straightforward. What was striking was the further philosophical conclusions Davidson inferred from these considerations. Far from deriving some sort of dualism, he used them to argue for a materialist monism. His argument is well known, but it bears repeating. Mental events, Davidson observed, enter into causal relations with physical events.⁷ But causal relations must be backed by laws; that is, causal relations between individual events must instantiate lawful regularities. Since there are no laws about the mental, either psychophysical or purely psychological, any causal relation involving a mental event must instantiate a physical law, from which it follows that the mental event has a physical description, or falls under a physical event kind. From this it further follows that the event is a physical event. For an event is physical (or mental) if it falls under a physical event kind (or a mental event kind).

It follows then that all events are physical events—on the assumption that every event enters into at least one causal relation. This assumption seems entirely unproblematic, for it only leaves out events that are both *causeless* and *effectless*. If there are any such events, it is difficult to see how their existence can be known to us; I believe we could safely ignore them. So imagine a Davidsonian universe of events: all these events are physical events, and some of them are also mental. That is to say, all events have physical properties, and some have mental properties as well. Such is Davidson's celebrated "anomalous monism".

Davidson's ontology recognizes individual events as spatiotemporal particulars. And the principal structure over these events is causal structure; the network of causal relations that interconnect events is what gives intelligible structure to this universe of events. What role does mentality play, on Davidson's anomalous monism, in shaping this structure? The answer: None whatever.

For anomalous monism entails this: the very same network of causal relations would obtain in Davidson's world if you were to redistribute mental properties over its events any way you like; you would not disturb a single causal relation if you randomly and

6. The classic source on reduction is Ernest Nagel, *The Structure of Science* (New York: Harcourt, Brace & World, 1961), ch. 11.

7. Actually the argument can proceed with a weaker premise to the effect that mental events enter into causal relations, either with physical events or with other mental events.

arbitrarily reassigned mental properties to events, or even removed mentality entirely from the world. The fact is that under Davidson's anomalous monism, mentality does no causal work. Remember: on anomalous monism, events are causes or effects only as they instantiate physical laws, and this means that an event's mental properties make no causal difference. And to suppose that altering an event's mental properties would also alter its physical properties and thereby affect its causal relations is to suppose that psychophysical anomalism, a cardinal tenet of anomalous monism, is false.⁸

Anomalous monism, therefore, permits mental properties no causal role, not even in relation to other mental properties. What does no causal work does no explanatory work either; it may as well not be there—it's difficult to see how we could miss it if it weren't there at all. That there are in this world just these mental events with just these mental characteristics is something that makes no causal difference to anything. On anomalous monism, that an event falls under a given mental kind is a causally irrelevant fact; it is also something that is entirely inexplicable in causal terms. Given all this, it's difficult to see what point there is in recognizing mentality as a feature of the world. I believe that if we push anomalous monism this way, we will find that it is a doctrine virtually indistinguishable from outright eliminativism.

Thus, what we see is this: anomalous monism, rather than giving us a form of nonreductive physicalism, is essentially a form of eliminativism. Unlike eliminativism, it allows mentality to exist; but mentality is given no useful work and its occurrence is left wholly mysterious and causally inexplicable. This doesn't strike me as a form of existence worth having. In this respect, anomalous monism does rather poorly even in comparison with epiphenomenalism as a realism about the mental. Epiphenomenalism gives the mental a place in the causal network of events; the mind is given a well-defined place, if not an active role, in the causal structure of the world.

These observations highlight the importance of *properties*; for it is in terms of properties and their interrelations that we make sense of certain concepts that are crucial in this context, such as law, causality, explanation, and dependence. Thus, the anomalousness of mental properties has far-reaching consequences within Davidson's framework: within it, anomalous properties are causally and explanatorily impotent, and it is doubtful that they can have any useful role at all. The upshot is that we don't get in Davidson's

8. Davidson says in "Mental Events" that he believes in the "supervenience" of the mental on the physical, and this does introduce a constraint on the distribution of physical properties when the distribution of mental properties is altered. This, however, does not detract substantively from the point being made here. For one, it remains true, on the notion of supervenience Davidson favors (which corresponds to "weak supervenience"; see his "Reply to Essays X-XII" in *Essays on Davidson: Actions and Events*, ed. Bruce Vermazen and Merrill B. Hintikka (Oxford: Oxford University Press, 1985)), that the removal of *all* mental properties from events of this world would have no consequence whatever on how physical properties are distributed over them. For another, the supervenience of the mental is best regarded as an independent thesis, and my present remarks only concern the implications of anomalous monism. I consider the supervenience view below in IV.

anomalous monism a plausible form of nonreductive physicalism; his anomalous monism comes perilously close to eliminativism.⁹

III

Let us now turn to the multiple realizability (or "compositional plasticity") of psychological events and its implications for psychophysical reduction. In a passage that turned out to have a profound impact on the discussions of the mind-body problem, Putnam wrote:¹⁰

Consider what the brain-state theorist has to do to make good his claims. He has to specify a physical-chemical state such that *any* organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusc's brain (octopuses are mollusca, and certainly feel pain), etc. At the same time, it must *not* be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extraterrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may *be* pain.

This paragraph helped bring on an unexpectedly early demise of the psychoneural identity theory of Smart and Feigl and inspired a new theory of the mental, functionalism, which in spite of its assorted difficulties is still the most influential position on the nature of the mental.¹¹ Putnam's basic point is that any psychological event-type can be "physically realized" or "instantiated" or "implemented" in endlessly diverse ways, depending on the physical-biological nature of the organism or system involved, and that this makes it

9. Davidson's overall views of the mental are richer and more complex than the present discussion might seem to indicate. I believe that they contain some distinctly dualistic elements; for a discussion of this aspect of Davidson, see my "Psychophysical Laws" in Ernest LePore and Brian McLaughlin, eds., *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Oxford: Blackwell, 1984). There have been some interesting recent attempts, which I cannot discuss here, to reconcile anomalous monism with the possibility of mental causation; see, e.g., Ernest LePore and Barry Loewer, "Mind Matters", *Journal of Philosophy* 84 (1987): 630-642; Brian McLaughlin, "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical", forthcoming; Terence Horgan, "Mental Quasation", forthcoming.

10. Putnam, "The Nature of Mental States".

11. Putnam himself has abandoned functionalism; see his *Representation and Reality* (Cambridge: MIT Press, 1988), chs. 5 and 6.

highly implausible to expect the event to correlate uniformly with, and thus be identifiable with, some "single" type of neural or physical state. This idea has been used by Fodor to formulate a general antireductionist argument, whose gist can be quickly summarized.

As we have seen, reduction of one theory to another is thought to require the derivation of the laws of the reduced theory from the laws of the reducer via "bridge laws". If a predicate of the theory being reduced has a nomologically coextensive predicate in the reducing theory, the universally quantified biconditional connecting the two predicates will be available for use as a bridge law.¹² Let us say that the vocabulary of the reduced theory is "strongly connected" with that of the reducing theory if such a biconditional bridge law correlates each predicate of the former with a predicate of the latter. It is clear that the condition of strong connectibility guarantees reduction (on the assumption that the theory being reduced is a true theory). For it would enable us to rewrite basic laws of the target theory in the vocabulary of the reducer, using these biconditional laws in effect as definitions. Either these rewrites are derivable from the laws of the reducing theory, or else they can be added as additional basic laws. In the latter case, the reducer theory has been expanded; but that does not diminish the ontological and conceptual import of the reductive procedure.

But what multiple realization puts in doubt, according to the antireductionist, is precisely the strong connectibility of mental predicates vis-à-vis physical-neural predicates. For any psychological property, there is in principle an endless sequence of nomologically possible physical states such that, though each of them "realizes" or "implements" it, none of them will by itself be coextensive with it. Why can't we take the *disjunction* of these physical states as the physical coextension of the mental property? Putnam somewhat disdainfully dismisses this move, saying only that "this does not have to be taken seriously".¹³ I think there are some complex issues here about disjunctive predicates vs. disjunctive properties, complexity of predicates vs. that of properties, etc.; but these are likely to be contentious issues that can only distract us at present.¹⁴ So let us go along with Putnam here and disregard the disjunctive solution to the multiple realization problem.

12. There are some complex logical and ontological details we are leaving out here. See, for details, Robert L. Causey, *Unity of Science* (Dordrecht: Reidel, 1977).

13. "The Nature of Mental States", p. 437.

14. Note also that derivational reduction does not *require* strong connectibility; any set of bridge laws, of whatever form and strength, will do as long as it enables the required derivation. But this obviously depends on the strength of the two theories involved, and there seems to be little of interest that is sufficiently general to say about this. There are also philosophical considerations for thinking that biconditionals and attribute identities are important in reduction. Cf. Lawrence Sklar, "Types of Inter-Theoretic Reduction", *British Journal for the Philosophy of Science* 18 (1967): 109-124.

In rejecting the disjunction move, however, Putnam appears to be assuming this: *a physical state that realizes a mental event is at least nomologically sufficient for it*. For if this assumption were rejected, the disjunction move couldn't even get started. This generates laws of the form " $P_i \rightarrow M$ ", where M is a mental state and P_i is a physical state that realizes it. Thus, where there is multiple realization, there must be psychophysical laws, each specifying a physical state as nomologically sufficient for the given mental state. Moreover, Putnam's choice of examples in the quotation above, which are either biological species or determinate types of physical structures ("extraterrestrials"), and his talk of "species-specificity" and "species-independence"¹⁵ suggest that he is thinking of laws of a somewhat stronger form, " $S_i \rightarrow (M \leftrightarrow P_i)$ ", which, *relative to species or structure* S_i , specifies a physical state, P_i , as *both necessary and sufficient* for the occurrence of mental state M . A law of this form states that any organism or system, belonging to a certain species, is such that it has the given mental property at a time if and only if it is in a certain specified physical state at that time. We may call laws of this form "species-specific biconditional laws."

In order to generate laws of this kind, biological species may turn out to be too wide; individual differences in the localization of psychological functions in the brain are well known. Moreover, given the phenomena of learning and maturation, injuries to the brain, and the like, the neural structure that subserves a psychological state or function may change for an individual over its lifetime. What is important then is that these laws are relative to physical-biological structure-types, although for simplicity I will continue to put the matter in terms of species. The substantive theoretical assumption here is the belief that for each psychological state there are physical-biological structure types, at a certain level of description or specification, that generate laws of this form. I think an assumption of this kind is made by most philosophers who speak of multiple realizations of psychological states, and it is in fact a plausible assumption for a physicalist to make.¹⁶ Moreover, such an assumption seems essential to the very idea of a physical realization; what else could "physical realization" mean?

So what I am saying is this: the multiple realization argument perhaps shows that the strong connectibility of mental properties vis-à-vis physical properties does not obtain: however, it *presupposes* that *species-specific strong connectibility* does hold. Merely to defeat the antireductionist argument, I need not make this second claim; all I need is the weaker claim that the phenomenon of multiple realization is *consistent* with the species-specific strong connectibility, and it seems to me that that is plainly true.

The point of all this is that the availability of species-specific biconditional laws linking the mental with the physical breathes new life into psychophysical reductionism. Unlike

15. "The Nature of Mental States", p. 437.

16. Ned Block says, "Most functionalists are willing to allow . . . that for each type of pain-feeling organism, there is (perhaps) a single type of physical state that realizes pain in that type of organism", in his "Introduction: What is Functionalism?" in Block, ed., *Readings in Philosophy of Psychology*, vol. 1 (Cambridge: Harvard University Press, 1980), p. 172. Such a law would have exactly the form under discussion.

species-independent laws, these laws cannot buy us a *uniform* or *global* reduction of psychology, a reduction of every psychological state to a uniform physical-biological base across all actual and possible organisms; however, these laws will buy us a series of *species-specific* or *local* reductions. If we had a law of this form for each psychological state-type for humans, we would have a physical reduction of human psychology; this reduction would tell us how human psychology is physically implemented, how the causal connections between our psychological events and processes work at the physical-biological level, what biological subsystems subserve our cognitive capacities and functions, and so forth. This is reduction in a full-blown sense, except that it is limited to individuals sharing a certain physical-biological structure. I believe "local reductions" of this sort are the rule rather than the exception in all of science, not just in psychology.¹⁷ In any case, this is a plausible picture of what in fact goes on in neurobiology, physiological psychology and cognitive neuroscience. And it seems to me that any robust physicalist must expect, and demand, the possibility of local reductions of psychology just in this sense.¹⁸

Thus, the conclusion we must draw is that the multiple realizability of the mental has no antireductionist implications of great significance; on the contrary, it entails, or at least is consistent with, the local reducibility of psychology, local relative to species or physical structure-types. If psychological states are multiply realized, that only means that we shall have multiple local reductions of psychology. The multiple realization argument, if it works, shows that a global reduction is not in the offing; however, local reductions are reduction enough, by any reasonable scientific standards and in their philosophical implications.

IV

Some have looked to the idea of "supervenience" for a formulation of physicalism that is free of reductionist commitments. The promise of supervenience in this area appears to have been based, at least in part, on the historical circumstance that some prominent ethical theorists, such as G.E. Moore and R.M. Hare, who constructed classic arguments against naturalistic reductionism in ethics, at the same time held the view that moral properties are "supervenient" upon descriptive or naturalistic properties. So why not think of the relation between psychological and physical properties in analogy with the relation, as conceived by these ethical theorists, between moral and descriptive properties? In each instance, the supervenient properties are in some substantive sense dependent on, or determined by, their subvenient, base properties and yet, it is hoped, irreducible to them.

17. See on this point Berent Enc, "In Defense of the Identity Theory", *Journal of Philosophy* 80 (1983): 279-98.

18. This point, and some related points, are elaborated in my "Disunity of Psychology as a Working Hypothesis?", forthcoming.

This was precisely the line of thinking that appears to have prompted Davidson to inject supervenience into the discussion of the mind-body problem. He wrote:¹⁹

Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respects, or that an object cannot alter in some mental respect without altering in some physical respect. Dependence or supervenience of this kind does not entail reducibility through law or definition: if it did, we could reduce moral properties to descriptive, and this there is good reason to *believe* cannot be done . . .

Although Davidson himself did not pursue this idea further, many other philosophers have tried to work this suggestive idea into a viable form of nonreductive materialism.

The central problem in implementing Davidson's suggestion has been that of defining a supervenience relation that will fill the twin requirements he set forth: first, the relation must be *nonreductive*; that is, a given domain can be supervenient on another without being reducible to it. Second, the relation must be one of *dependence*: if a domain supervenes on another, there must be a sturdy sense in which the first is dependent on the second, or the second determines the first. But it has not been easy to find such a relation. The main difficulty has been this: if a relation is weak enough to be nonreductive, it tends to be too weak to serve as a dependence relation; conversely, when a relation is strong enough to give us dependence, it tends to be too strong—strong enough to imply reducibility.

I will not rehearse here the well known arguments pro and con concerning various supervenience relations that have been proposed. I will instead focus on one supervenience relation that has seemed to several philosophers²⁰

1y to hold the most promise as a nonreductive dependency relation, viz., "global supervenience". The generic idea of supervenience is that things that are indiscernible in respect of the "base" (or "subvenient") properties cannot differ in respect of the supervenient

19. "Mental Events", in Davidson, *Essays on Actions and Events*, p. 214.

20. Including Terence Horgan in his "Supervenience and Microphysics", *Pacific Philosophical Quarterly* 63 (1982): 29-43; John Haugeland in "Weak Supervenience", *American Philosophical Quarterly* 19 (1982): 93-103; John Post in *The Faces of Existence* (Ithaca: Cornell University Press, 1987); and Bradford Petrie, "Global Supervenience and Reduction", *Philosophy and Phenomenological Research* 48 (1987): 119-130. The model-theoretic notion of determination worked out by Geoffrey Hellman and Frank Thompson, in "Physicalism: Ontology, Determination, and Reduction", *Journal of Philosophy* 72 (1975): 551-564, is closely related to global supervenience.

properties. Global supervenience applies this consideration to "worlds", giving us the following formulation of psychophysical supervenience:

Worlds that are indiscernible in all physical respects are indiscernible in mental respects; in fact, physically indiscernible worlds are one and the same world.

Thus, any world that is just like this world in all physical details must be just like it in all psychological respects as well. This relation of supervenience is appropriately called "global" in that worlds rather than individuals within worlds are compared for discernibility or indiscernibility in regard to sets of properties. What is it for two worlds to be physically, or mentally, indiscernible? For simplicity let us assume that the same individuals exist in all the worlds.²¹ We may then say that two worlds are indiscernible with respect to a set of properties just in case these properties are distributed over individuals in the same way in the two worlds.

It can be shown that, as hoped, the global supervenience of the mental on the physical does not entail the existence of psychophysical laws;²² thus, global supervenience is consistent with the nomological irreducibility of the mental to the physical. The only question then is whether it yields an appropriate relation of dependency between the mental and the physical, one that is strong enough to qualify it as a physicalism. The answer, I will argue, is in the negative.

We may begin by observing that the global supervenience of the mental permits the following: Imagine a world that differs from the actual world in some minute physical detail. We may suppose that in that world one lone hydrogen atom somewhere in deep space is slightly displaced relative to its position in this world. This world with one wayward hydrogen atom could, consistently with the global supervenience of the mental, be as different as you please from the actual world in any mental respect (thus, in that world nothing manifests mentality, or mentality is radically redistributed in other ways). The existence of such a world and other similarly aberrant worlds does not violate the constraints of global supervenience; since they are not physically indiscernible from the actual world, they could, under global supervenience, differ radically from this world in psychological characteristics.²³

21. Even with this simplifying assumption certain complications arise; however, we may disregard them for the present purposes. For further details see my "Supervenience for Multiple Domains", *Philosophical Topics* 16 (1988): 129-150.

22. At least not in a straightforward way. See my "'Strong' and 'Global' Supervenience Revisited", *Philosophy and Phenomenological Research* 48 (1987): 315-326.

23. This particular difficulty can be largely met by formulating global supervenience in terms of *similarity* between worlds rather than indiscernibility. See my "'Strong' and 'Global' Supervenience Revisited".

If that doesn't convince you of the weakness of global supervenience as a determination or dependency relation, consider this: it is consistent with global supervenience for there to be two organisms in our actual world which, though wholly indiscernible physically, are radically different in mental respects (say, your molecule-for-molecule duplicate is totally lacking in mentality). This is consistent with global supervenience because there might be no other possible world that is just like this one physically and yet differing in some mental respect.²⁴

It seems to me that indiscernibility considerations at the global level, involving whole worlds, are just too coarse to give us the kind of dependency relation we should demand if the mental is truly dependent on the physical. Like it or not, we treat individuals, and perhaps also aggregates of individuals smaller than total worlds, as psychological units, and it seems to me that if psychophysical determination or dependence means anything, it ought to mean that the psychological nature of each such unit is wholly determined by its physical nature. That is, dependency or determination must hold at the local as well as the global level.

Moreover, talk of whole worlds in this connection, unless it is anchored in determinative relations obtaining at the local level, has little verifiable content; it is difficult to see how there can be empirical evidence for the global supervenience thesis that is not based in evidence about specific psychophysical dependencies—dependencies and correlations between specific psychological and physical properties. In fact, it seems to me that we must look to local dependencies for an *explanation* of global supervenience as well as its evidence. Why is it the case that no two worlds can exist that are indiscernible physically and yet discernible psychologically? Or why is it the case that “physical truths determine all the truths,”²⁵ as some prefer to put it? I think this is a legitimate question to raise, and as far as I can see the only answer, other than the response that it is a brute, unexplainable metaphysical fact, is in terms of local correlations and dependencies between specific mental and physical properties. If the global supervenience of the mental on the physical were to be proposed as an unexplainable fact that we must accept on faith, I doubt that we need to take the proposal seriously. Specific psychophysical dependencies holding for individuals, and other proper parts of the world, are both evidence for, and an explanatory ground of, global supervenience.

The trouble is that once we begin talking about correlations and dependencies between specific psychological and physical properties, we are in effect talking about psychophysical laws, and these laws raise the specter of unwanted physical reductionism. Where there are psychophysical laws, there is always the threat, or promise, of psychophysical reduction. We must conclude that supervenience is not going to deliver to us a viable form of nonreductive materialism.

24. This shows that global supervenience is consistent with the failure of “weak supervenience”. See my “‘Strong’ and ‘Global’ Supervenience Revisited”.

25. See Hellman and Thompson, “Physicalism: Ontology, Determination, and Reduction”; Post, *The Faces of Existence*.

V

So far I have reviewed three influential formulations of nonreductive materialism—Davidson's anomalous monism, the Putnam-Fodor doctrine of psychological autonomy, and supervenient physicalism—and found each of them wanting either as a materialism or as an antireductionism. In this final section, I want to advance a direct argument to show why the prospects for a nonreductive physicalism are dim.

Let us first of all note that nonreductive physicalism is not to be a form of eliminativism; that is, it acknowledges the mental as a legitimate domain of entities. What sort of entities? Here let us, for convenience, make use of the Davidsonian scheme of individual events, thinking of mentality to be exhibited as properties of these events. Thus, as a noneliminativist, the nonreductive physicalist believes that there are events in her ontology that have mental properties (e.g., being a pain, being a belief that snow is cold, etc.). I argued earlier, in discussing Davidson's anomalous monism, that if your noneliminativism is to be more than a token gesture, you had better find some real causal work for your mental properties. The fact that a given event is a mental event of a certain kind must play some causal-explanatory role in what other events occur and what properties they have. Thus, I am supposing that a nonreductive physicalist is a mental realist, and that to be a mental realist, your mental properties must be *causal properties*—properties in virtue of which an event enters into causal relations it would otherwise not have entered into.

Let me now make another assumption: psychophysical causation takes place—that is, some mental events cause physical events. For example, a sudden sharp pain felt in my hand causes a jerky withdrawal of the hand. It is true that in a Davidsonian domain, all events are physical; that is, every event has some physical property. But when I say that mental events cause physical events, something stronger is intended, namely that an event, *in virtue of its mental property*, causes another event to have a certain physical property. I believe that this assumption will be granted by most of us—it will be granted by anyone who believes that at least sometimes our limbs move because we have certain desires and beliefs.²⁶ When I walk to the water fountain for a drink of water, my legs move in the way they do in part because of my desire for water and my belief that there is water to be had at the water fountain.

There is a further assumption that I believe any physicalist would grant. I call this “the causal closure of the physical domain”; roughly, it says this: *any physical event that has a cause at time t has a physical cause at t* . This is the assumption that if we trace the causal ancestry of a physical event, we need never go outside the physical domain. To deny this assumption is to accept the Cartesian idea that some physical events have only nonphysical causes, and if this is true there can in principle be no complete and self-sufficient physical theory of the physical domain. If the causal closure failed, our physics would need to refer in an essential way to nonphysical causal agents, perhaps Cartesian

26. For a forceful statement of this point see Fred Dretske, *Explaining Behavior: Reasons in a World of Causes* (Cambridge: MIT Press, 1988).

souls and their psychic properties, if it is to give a complete account of the physical world. I think most physicalists would find that picture unacceptable.

Now we are ready to derive some consequences from these assumptions. Suppose that a certain event, in virtue of its mental property, causes a physical event. The causal closure of the physical domain says that this physical event must also have a physical cause. We may assume that this physical cause, in virtue of its physical property, causes the physical event. The following question arises: *What is the relationship between these two causes, one mental and the other physical?* Each is claimed to be a cause of the physical effect. There are two initial possibilities that we can consider.

First, when we are faced with two purported causes of a single event, we could entertain the possibility that each is only a *partial cause*, the two together making up a full or sufficient cause, as when a car crash is said to be caused by the driver's careless braking and the icy condition of the road. Applied to our case, it says that the mental cause and the physical cause are each only a partial cause, and that they *together* make up one sufficient cause. This seems like an absurd thing to say, and in any case it violates the causal closure principle in that it regards the mental event as a necessary constituent of a full cause of a physical event; thus, on this view, a full causal story of how this physical event occurs must, at least partially, go outside the physical domain.

Could it be that the mental cause and the physical cause are each an *independent sufficient* cause of the physical effect? The suggestion then is that the physical effect is *overdetermined*. So if the physical cause hadn't occurred, the mental cause by itself would have caused the effect. This picture is again absurd: from what we know about the physiology of limb movement, we must believe that if the pain sensation causes my hand to withdraw, the causal chain from the pain to the limb motion must somehow make use of the causal chain from an appropriate central neural event to the muscle contraction; it makes no sense to think that there was an independent, perhaps telekinetic, causal path from the pain to the limb movement. Moreover, the overdetermination idea seems to violate the causal closure principle as well: in the counterfactual situation in which the physical cause does not occur, the closure principle is violated. For the idea that the mental and the physical cause are each an independent sufficient cause involves the acceptance of the counterfactual that if the physical cause had not occurred, the mental cause would have occurred and caused the physical effect. This is in violation of the causal closure principle.

These two ways of looking at the situation are obvious nonstarters. We need a more plausible answer to the question, how are the mental cause and the physical cause of the single physical effect related to each other? Given that any physical event has a physical cause, how is a mental cause *also* possible? This I call "the problem of causal-explanatory exclusion", for the problem seems to arise from the fact that a cause, or causal explanation, of an event, when it is regarded as a full, sufficient cause or explanation, appears to *exclude* other *independent* purported causes or causal explanations of it.²⁷

27. This idea is developed in greater detail in my "Mechanism, Purpose, and Explanatory Exclusion", *Philosophical Perspectives* 3 (1989), forthcoming.

At this point, you might want to protest: why all this beating around the bush? Why not just say the mental cause and the physical cause are one and the same? Identification simplifies ontology and gets rid of unwanted puzzles. Consider saying that there are in this glass two distinct substances, H_2O and water; that is, consider saying that water and H_2O co-occur everywhere as a matter of law but that they are distinct substances nonetheless. This would invite a host of unwanted and unnecessary puzzles: given that what is in the glass weighs a total of ten ounces, how much of the weight is to be attributed to the water and how much to the H_2O ? By dropping a lighted match in the glass, I extinguish it. What caused it? Was it the water or the H_2O ? Were they each only a partial cause, or was the extinguishing of the match overdetermined? The identification of the water with the H_2O puts all these questions to rest in a single stroke: there is here one thing, not two. The identity solution can work similar magic in our present case: the pain is a neural state—here there is one cause, not two. The limb motion was caused by the pain, that is to say, by a neural state. The unwanted puzzles vanish.

All this is correct. But what does the identity solution involve? Remember that what is for us at issue is the causal efficacy of *mental properties* of events vis-à-vis their physical properties. Thus, the items that need to be identified are properties—that is, we would need to identify mental properties with physical properties. If this could be done, that would be an excellent way of vindicating the causal powers of mentality.

But this is precisely the route that is barred to our nonreductivist friends. The identification of mental properties with physical properties is the heart of reductionist “type physicalism”. These property identities would serve as bridge laws par excellence, enabling a derivational reduction of psychology to physical theory. The identities entail psychophysical correlations of biconditional form, stable over possible, or nomologically possible, worlds, and this, we have been told, is excluded by Davidson’s mental anomalism and Putnam’s multiple realization argument. So the identity solution is out of the question for the nonreductive materialist. Is there any other way to respond to the causal exclusion problem, a way that falls short of identifying mental with physical attributes?

There is one, but it isn’t something that would be palatable to the nonreductivist. I believe that the only way other than the identity solution is to give a general account of causal relations involving macro-events as “supervenient causal relations”, causal relations that are supervenient on micro-causal processes. You put a kettle of water on the stove and turn on the burner; and soon the water starts to boil. Heating the water caused it to boil. That is a causal relation at the macro-level. It is natural to think of this causal relation as supervenient on certain underlying causal processes at the micro-level. The heating of water supervenes on the increasing kinetic energy of water molecules, and when their mean kinetic energy reaches a certain level, water molecules begin to move in turbulence, some of them being ejected into the air. Boiling is a macro-state that supervenes on just these micro-processes. A sharp pain causes an anxiety attack five seconds later. What’s going on? Again, it is tempting, and natural, to think thus: the pain is supervenient on a certain underlying neural activity, and this neural event causes another neural event to occur. The anxiety attack occurs because it is supervenient on this second neural event.

The general model of supervenient causation applied to macro-causal relations is this: macro-event **m** is a cause or effect of event **E** in virtue of the fact that **m** is supervenient

on some micro-event, *n*, which is a cause or effect of event *E*.²⁸ The suggestion then is that we use this model to explain mental causation: a mental event is a cause, or an effect, of another event in virtue of the fact that it is supervenient on some physical event standing in an appropriate causal relation to this event. Thus, mental properties are seen as deriving their causal potential from the physical properties on which they supervene. That is the main idea.

But what sort of supervenience relation is involved in this picture? Global supervenience we considered above obviously will not do; it does not give us a way of speaking of supervenience of specific mental properties on specific physical properties, since it only refers to indiscernibility holding for worlds. Supervenient causation in my sense requires talk of specific mental properties supervening on specific physical base properties, and this is possible only if there are laws correlating psychological with physical properties. This is what I have called elsewhere "strong supervenience", and it can be argued plausibly that supervenience of this strength entails the possibility of reducing the supervenient to the subvenient.²⁹ I will spare you the details here, but the fact that this form of supervenience directly involves psychophysical laws would be enough to give pause to any would-be nonreductive physicalist. I am not entirely certain that this supervenience solution will suffice; that is, I am not certain that anything short of the identity solution will resolve the exclusion problem. However, I believe that it is the only alternative to explore if, for whatever reason, you are unwilling or unable to go for psychophysical attribute identities. But I doubt that this solution will be found acceptable by the nonreductivist any more than the identity solution.

28. For critical discussions of this model, see Brian McLaughlin, "Event Supervenience and Supervenient Causation", *Southern Journal of Philosophy* 22, *The Spindel Conference Supplement on Supervenience* (1984): 71-91; Peter Menzies, "Against Causal Reductionism", *Mind* 97 (1988): 560-574.

29. I am putting the point somewhat tentatively here because it involves several currently contentious issues. For a general argument for this point, see my "Concepts of Supervenience", *Philosophy and Phenomenological Research* 45 (1984): 153-176; especially, section III; and "Supervenience as a Philosophical Concept", forthcoming in *Metaphilosophy*. However, this argument makes use of infinite disjunctions and conjunctions (actually, infinite disjunctions are all one needs; see "Supervenience as a Philosophical Concept"). If the argument is found objectionable because of this feature, it could be supplemented with an argument modeled on my argument in section III above against the Putnam-Fodor antireductionist thesis. This means that the supervenience relation needed for the model of supervenient causation sketched here must require that each supervenient property have a *nomologically coextensive base property relative to the given physical structure*. There are, I believe, plausible considerations in favor of this stronger supervenience relation as a basis for the concept of supervenient causation (or the reduction of causal relations); however, I cannot go into the details here.

If nonreductive physicalists accept the causal closure of the physical domain, therefore, they have no visible way of accounting for the possibility of psychophysical causation. This means that they must either give up their antireductionism or else reject the possibility of psychophysical causal relations. The denial of psychophysical causation can come about in two ways: first, you make such a denial because you don't believe there are mental events; or second, you keep faith with mental events even though you acknowledge that they never enter into causal transactions with physical processes, constituting their own autonomous causal world. So either you have espoused eliminativism, or else you are moving further in the direction of dualism, a dualism that posits a realm of the mental in total causal isolation from the physical realm. This doesn't look to me much like materialism.

Is the abandonment of the causal closure of the physical domain an option for the materialist? I think not: to reject the closure principle is to embrace irreducible nonphysical causes of physical phenomena. It would be a retrogression to Cartesian interactionist dualism, something that is definitive of the *denial* of materialism.

Our conclusion, therefore, has to be this: nonreductive materialism is not a stable position. There are pressures of various sorts that push it either in the direction of outright eliminativism or in the direction of an explicit form of dualism.³⁰

30. My thanks to Richard Brandt, Sydney Shoemaker, and Ernest Sosa for helpful comments on earlier versions, and to David Benfield, Barry Loewer, and Brian McLaughlin for discussing with me some of the topics of this paper.