

# 16 *Mind–Body Causation and Explanatory Practice*

In recent years a number of philosophers have worried about whether we can reasonably believe that mental properties are causally efficacious. They are concerned whether the intentional ‘aspects’ or qualitative ‘aspects’ of mental events are epiphenomenal—that is, lacking in causal power and irrelevant to causal transactions. They typically assume that mental events themselves are causes. But this is supposed to be ensured by the prior assumption that mental events are physical events. Individual mental events are assumed to be instantiations or tokens of physical event-kinds. The doubt is whether the mental ‘aspects’ of these states and events play any significant role in causal processes.

I think that these worries can be met within the materialist metaphysical framework in which they arise. I will say a little about what I think is wrong with some of the more prominent sorts of argument that lead to epiphenomenalism. But what interests me more is the very existence of the worries. I think that they are symptomatic of a mistaken set of philosophical priorities. Materialist metaphysics has been given more weight than it deserves. Reflection on explanatory practice has been given too little. The metaphysical grounds that support the worries are vastly less strong than the more ordinary grounds we already have for rejecting them.

I shall first outline the worries about epiphenomenalism, and identify some of the weak spots in the arguments for them. Then I shall explain why I think that the starting point for the worries is itself dubious. Finally I shall generalize a little about the misguided priorities (as I see them) that engender these worries.

1

The picture that leads to the worries about epiphenomenalism begins with a plausible idea. It is that certain non-intentionally described states or events that ‘underlie’ mental states and events participate in causal processes that are instances of physical laws that do not mention mentality. I want to leave open

what ‘underlie’ is to mean here. I will assume, however, at least that mental states and events would not occur if some ‘underlying’ physical states and events did not occur. There are no gaps in these physical chains of events. So, for example, there are underlying, gapless neural processes that are instances of laws of neurophysiology; and the mental events would not occur if some such processes did not occur. I have no serious doubts about this view.

These physical states and events are usually thought to ‘underlie’ the mental states and events in a more specific sense. They are held to be token-identical with them. I do not accept this materialist claim. But it is widely accepted, largely because of its supposed virtues in clarifying mental causation. From the time of Descartes it has often been thought that there is some mystery in how mental and physical events can interact. This and other materialist views purport to dispel the mystery by holding that there is only one sort of causation—a relation between physical events. In this section, I will accept this ontology, though only for the sake of argument.

The worry about epiphenomenalism arises by considering that some properties of events (or states) are relevant to their causal relations, while others are not. The property of being the third large explosion in a given country during the last ten years may be irrelevant to that explosion’s causing certain damage, whereas the heat of the explosion would be relevant. The question is whether mentalistic properties are causally efficacious, or whether mentalistic descriptions of properties are relevant to understanding causation. Many philosophers have developed more confidence in the causal efficacy of the underlying neural properties than in that of the ‘intentional aspects’ of mental events. If those neural processes are going on, the body’s movements, and hence what we count as behavior, will depend on the properties of those processes. The intentional or phenomenal ‘aspects’ of the mental events might, they think, be irrelevant or at best quite derivative and indirect—epiphenomenal on the real underlying causal processes.

A comparison of mental properties to properties like phenotypes in biology is typical. There are regular, even—assuming richly filled-in background conditions—loosely nomological, relations between the phenotypes of parents and phenotypes of their immediate offspring. But the phenotypes of the parents are causally inefficacious in producing those of their offspring. The real causal efficacy derives from the parental genotypes. Some philosophers appear to be seriously concerned that intentional kinds are, like phenotypes, part of a nomologically describable system, but not causally efficacious in their own right.

What motivates these worries? Broadly and crudely speaking, it is the picture that the underlying processes are occurring anyway and that the mental events really derive their causal efficacy from the physical properties which are the real agents of causation. This picture leads to the idea that the mental properties are superfluous. Even if the mental properties were not separable from the physical properties (because they necessarily supervened on them), they would be in a sense along for the ride, since the primary mechanisms of causation are located in the underlying physical properties.

Let us cast this sort of motivation into an argument. One could assume: (A) that mental event-tokens are identical with physical event-tokens; (B) that the causal powers of a physical event are determined only by its physical properties; and (C) that mental properties are not reducible to physical properties. From these assumptions, it may seem to follow that a mental event's mental properties play no role in determining its causal powers.

Here is another argument for the picture: Assume (a) that the world of physical events and properties is a complete and closed system, in the sense that physical events can be caused only by virtue of physical properties of other physical events; (b) that mental properties are not reducible to physical properties. It may again seem to follow that no physical events can be caused by virtue of mental properties, even if these mental properties are properties of physical events. The argument does leave open the possibility that mental properties of mental events are causally efficacious with respect to other mental properties of mental events. But if mental causation has no outlet among physical events, it is surely a peculiarly limited sort of causation.<sup>1</sup>

Let us begin with the first argument. As is common in arguments for epiphenomenalism, the key phrases leave much clarity to be desired. 'Is determined by' is a case in point. It could mean 'supervenes on', 'is explained by', or 'is individuated by'. For reasons that will emerge, I need not choose among these readings. (B) is problematic on all of them.

It is also not clear what is to be included in the notion of a physical property. In particular, it is unclear whether various relations to the environment count among the physical 'properties'. If they do not, the premiss is difficult to defend. I will assume that a very broad notion of physical property is intended—one that encompasses relations to the environment which are described in non-mentalist, non-intentional terms.

The fundamental unclarity lies in the notion of causal power.<sup>2</sup> The causal powers of a kind of event are to be understood in terms of the patterns of causation that events of that kind enter into. Such patterns are identified as explanatory in causal explanations. And the properties that 'determine' the causal powers of an event are those that enter into causal explanations. The second premise is plausible only insofar as one considers the causal powers of a physical event to be got only through patterns of properties described in the physical sciences, or in other commonsense explanations in physical terms. The sense in which only physical properties determine the causal powers of a physical event is just that within the patterns of causation described in the physical sciences and commonsense physicalistic discourse, physical properties suffice to provide a basis for the

<sup>1</sup> For a discussion of such arguments with which I have some sympathy, see Robert Van Gulick, "Who's in Charge Here? And Who's Doing All the Work?", in J. Heil and A. Mele (eds.), *Mental Causation* (Oxford: Oxford University Press, 1993), ch. 13. My criticisms of these arguments differ from his, but are for the most part compatible.

<sup>2</sup> Cf. my "Individualism and Causation in Psychology", *Pacific Philosophical Quarterly*, 70 (1989), 303–322, (Ch. 14 above) for a detailed account of my understanding of causal power.

existence and understanding of the causal powers of physical events; and no other properties enter in. Both the chains of causation and the patterns of explanation are in no need of supplementation from outside the realm of physical properties or physicalistic discourse. This is a tempting and plausible interpretation of the second premise. But on this interpretation the conclusion will not follow.

If physical events have mental properties, one is not entitled to the view that only physical properties (properties specified in the physical sciences or in ordinary physicalistic discourse) determine all the causal powers of a physical event (as opposed to merely all the causal powers associated with physicalistic explanations of the physical event), unless one can show that mentalistic explanation is either non-causal or fails to describe patterns of causal properties. For the causal powers of a physical event that is mental might include possible effects that are specified in mentalistic explanation. No one has shown that mentalistic explanation is either non-causal or non-descriptive. Nor is either view plausible.

Normally we consider an entity's causal powers relative to the kind in terms of which the entity is specified. For example, in asking for the causal powers of the heart, we implicitly expect physiological patterns of properties to be cited. We do not expect citation of powers that would be studied by physics. Pumping blood is usually considered relevant; squashing a bug if dropped from a ladder is not. If we were to specify the heart as a physical object of such and such physical dimensions, the latter property would seem relevant. The second premise (*B*) of the argument attracts an interpretation that is plausible but insufficient for the argument because the mental events are specified as physical. Then mental properties seem irrelevant to its causal powers. But if it is specified as mental ('What are the causal powers of a thought that it is raining?'), the idea that only properties specified in the physical sciences are relevant to determining the causal powers seems outlandish. Thus one cannot just take the second assumption of the argument as a generalized self-evident metaphysical principle.<sup>3</sup> Interpreted in a way that leaves the argument valid, the premise is either false or question-begging.

<sup>3</sup> Sometimes the second premise is stated, 'The causal powers of a physical event are *completely* determined by its physical properties.' This premise has substantially the same difficulties as the one I discuss. This premise is, however, weaker in that it allows the possibility that although the physical properties completely determine the causal powers of a physical event, mental properties may also play a role in determining those powers. To complete the argument, one needs a premise excluding 'overdetermination'. Issues about 'overdetermination' are fairly similar to those that I am discussing. I might say, however, that I find the term extremely misleading inasmuch as it assimilates different levels or ranges of causal interaction to ordinary cases in which (say) physical causes occur simultaneously and each is sufficient unto itself for their common effect. There are deep differences in the two cases.

I might add here that in all these arguments, a materialist must have some explanation for regarding mental properties as material in character. Their existence must not be seen as incompatible with the materialism. I think that this is not an easy problem, but there are various solutions that seem to satisfy many materialists. And since I am accepting materialism in this section for the sake of argument, I shall not pursue the matter.

Let us turn to the second argument. Problems with the second argument lie in the first premise, (a). The claim that physical events can be caused only by virtue of physical properties of other physical events has problems entirely analogous to those that beset the first argument. The existence of a closed system reflects a pattern of causal relations and of causal explanation that needs no supplementation from the outside. There are no gaps. It does not follow from this that such a system excludes or overrides causal relations or causal explanation in terms of properties from outside the system. Indeed, if it did follow, as has often been pointed out, there would be no room for causal efficacy in the special sciences, even in natural sciences like chemistry and physiology. For there is no gap (other than perhaps quantum gaps) in the causal relations explained in terms of the properties of physics. But few are tempted by the idea that physical events cannot be caused in virtue of physiological properties of physical events.

Is the causal efficacy of properties cited in chemistry and physiology dependent on the reducibility of the properties cited in these fields to those cited in physics? That seems almost equally outlandish. It is a wide-open empirical question whether properties of these special sciences are reducible to those cited in physics. In fact, it seems very unlikely that general reduction is possible. The causal relevance of the properties of these special sciences seems independent of questions of reducibility.

Weaknesses in the foregoing arguments are widely known. But there remains a sense of unease. Many seem disturbed by the picture of mental properties supervening on physical properties and just going along for the ride. Thus there have been various attempts to state what kind of supervenience relations would ‘allow’ mental properties to be causally efficacious even though they supervene on physical properties.<sup>a</sup>

These projects can be interesting. But in my view, the worries about epiphenomenalism have an air of make-believe. It is much surer that epiphenomenalism is false than that the various assumptions (even including the materialist assumptions) that have been thought to lead to it are true. It is also much surer that epiphenomenalism is false than that the various attempts to show it false or avoidable by appeal to counterfactuals, accounts of laws, or supervenience, are true. Epiphenomenalism is often taken as a serious metaphysical option. But it is better seen as at best a source of pressure for clarifying our common conceptions. It is rather like one of the less plausible scepticisms, which can be used as an instrument for philosophical clarification, but which has little real persuasive force. I think that a different, less metaphysical attitude in thinking about the problem would be more realistic and fruitful.

The irony is that by trying to clarify mentalistic causation, many materialists have come to believe that there is a serious metaphysical issue whether

<sup>a</sup> Jaegwon Kim, “Epiphenomenal and Supervenient Causation”, *Midwest Studies in Philosophy*, 9 (1984), 257–270; Ernest Sosa, “Mind–Body Interaction and Supervenient Causation”, *Midwest Studies in Philosophy*, 9 (1984), 271–281.

mentalist characterizations have any causal relevance at all. I think that this is tantamount to admitting that materialism has failed to illumine mental causation. But such illumination has been advertised as materialism's chief selling-point. Although materialism is not forced to accept epiphenomenalism, the very fact that the view is taken so seriously as a metaphysical option suggests that something has gone wrong in the search for clarification.

One cannot understand mentalistic causation (causation involving mentalistic or intentional properties) and mental causal powers by concentrating on properties characterized in the physical sciences. Our understanding of mental causation derives primarily from our understanding of mentalistic explanation, independently of our knowledge—or better, despite our ignorance—of the underlying processes. Materialist accounts have allowed too wide a gap between their metaphysics of mental causation and what we actually know about the nature and existence of mentalistic causation, which derives almost entirely from mentalistic explanations and observations.

2

So far I have not questioned the materialist metaphysics that helps ground the worries about epiphenomenalism. In this section I want to advance reasons for doubting the most common form of materialism and its centrality in understanding mental causation.

There is certainly reason to believe that underlying our mental states and processes are physical, chemical, biological, and neural processes that proceed according to their own laws. Some such physical processes are probably necessary if intentional (or phenomenal) mental events are to be causes of behavior. They seem necessary even for the mental events to exist. But, in my view, the nature of the relation between mental events and these physical processes is thoroughly unclear. The most widely accepted account of the relation is the materialist token-identity theory.

Some years ago I gave an argument against a significant version of the materialist token-identity theory.<sup>b</sup> The version I had in mind holds that each mental-state instance and event-token is identical with a physical-state instance or event-token that instantiates a physical natural kind specified in some actual natural science, or specifiable in some reasonable extension of the natural sciences as we now know them. Different mental event-tokens of the same mental event type may be tokens of different physical natural kinds.

The requirement that the physical event-token instantiate a physical kind specifiable in a natural science (physics, chemistry, biology, neurophysiology, and so on) is meant to ensure that the materialist utilize a non-question-begging

<sup>b</sup> Burge, "Individualism and the Mental" *Midwest Studies in Philosophy*, 4 (1979), 73–121 (Ch. 5 above).

identification that is not only uncontroversially physical, but plays some role in explanation of physical causation. I think that there are materialist views that are less committal than this one. My argument does not defeat these views. I shall remark on some of them later.

The argument against this sort of token-identity theory is partly based on twin earth thought experiments that I shall presume are familiar.<sup>4</sup> According to these thought experiments, it is possible for a person's body, considered in isolation from its relations to the environment, to be physiologically and molecularly the same even if the person were to think thoughts that have different intentional content. The difference in content depends on differences in the individual's historical relations to his or her environment. For example, it is possible for a person who has borne some historical relation (perhaps through vision or interlocution) to aluminum or arthritis to think that aluminum is a light metal or that arthritis is a painful disease, even though the person has no dispositions that would enable him to discriminate aluminum or arthritis from all other actual or possible metals or diseases—except by thinking of it as aluminum or arthritis. Counterfactual environments are possible in which one of these other 'look-alike' metals or diseases plays the same role in the acquisition and production of thoughts that aluminum and arthritis actually do. In such counterfactual environments the person might, for all intents and purposes, have a body that is a chemical and physiological duplicate of the actual person's body. Yet the person would be thinking thoughts with different content. The person would not be thinking that aluminum is a light metal or that arthritis is a painful disease.

I shall take it for granted that these thought experiments are sound, as so far described. The first premise of the argument against the token-identity theory is strongly suggested (though not entailed) by the thought experiments:

- (1) It is possible for a person to think thoughts with different contents even though all event-tokens that occur in the individual's body, that are plausible candidates for identification with mental events, and that are specifiable by physical sciences such as physics, chemistry, and neurophysiology, are the same.

The second premise is less specifically related to the thought experiments:

- (2) No occurrence of a thought could have a different intentional content and be the very same token-event or event-particular.

<sup>4</sup> *Ibid.*; Burge, "Other Bodies", in A. Woodfield (ed.), *Thought and Object: Essays on Intentionality* (Oxford: Clarendon Press, 1982), 97–120; *idem*, "Cartesian Error and the Objectivity of Perception", in P. Pettit and J. McDavell (eds.), *Subject, Thought, and Context* (Oxford: Clarendon Press, 1986), 117–136; *idem*, "Intellectual Norms and the Foundations of Mind", *The Journal of Philosophy*, 83 (1986), 697–720; *idem*, "Wherein is Language Social?", in A. George (ed.), *Reflections on Chomsky* (Oxford: Blackwell, 1989), 175–191 (Chs. 4, 7, 10, 11 above). The thought experiments use the methodology set out in Hilary Putnam, "The Meaning of 'Meaning'", in *Philosophical Papers*, ii (Cambridge: Cambridge University Press, 1975), 215–271.

Now take any physical event-token *b* in the individual's body that is a plausible candidate for being identical with the individual's occurrent thought (mental event-token) *a* that aluminum is a light metal (or that arthritis is a painful disease). By (1), there are possible situations in which the same token *b* occurs, but in which there occur only thoughts (mental event-tokens) with different intentional content. By (2), none of these thought occurrences is the very same token event as *a*. So since *b* could occur without *a*'s occurring, *b* cannot be identical with *a*.

This argument has been criticized by Donald Davidson. Davidson accepts the second premise, calling its denial 'not merely implausible but absurd. If two mental events have different contents, they are surely different events.' He does not squarely confront the first premise. But he appears to reject it. He thinks that the relevant thought experiments show that 'people who are in all relevant respects similar ... can differ in what they mean or think. ... But of course there is *something* different about them, even in the physical world; their causal histories are different.'<sup>5</sup>

I find this response unconvincing. There certainly are physical differences between actual and counterfactual situations in the relevant thought experiments. The question is whether there are always physically different entities that are plausible candidates for being identical with the different mental events or state-instances. The different physical causal histories are not plausible candidates. These histories do not have the same causes or effects that the relevant mental events (states) do. Moreover, it is doubtful that relevantly described causal histories instantiate explanatory natural kinds in any of the physical sciences.

One might think that since, in the counterfactual situation, there are differences at least in the remote causal ancestry of every relevant physical event in the individual's body, every such event would be a different event-token from any event-token in the actual situation. Events are, on this view, different just by virtue of having some difference in their causal histories.<sup>6</sup> But this seems an extremely implausible view of event identity. It seems to me clearly possible to consider the same event-token in an individual's body in a counterfactual situation, without being committed to the view that every event in the causal ancestry of that event, however remote, remains token- (or type-) identical. Much counterfactual reasoning about events depends on not being so inflexible. We frequently talk about a particular event under counterfactual suppositions in

<sup>5</sup> The quotations are from D. Davidson, "Knowing one's own Mind", *Proceedings of the American Philosophical Association*, 60 (1987), 452. Davidson has helpfully confirmed his rejection of premise (1) in private communication.

<sup>6</sup> This may be a consequence of Davidson's criterion for individuating events; cf. D. Davidson, "The Individuation of Events", in N. Rescher (ed.), *Essays in Honor of Carl G. Hempel* (Dordrecht: Reidel, 1969), 216–234. But since the criterion is not formulated modally, it is not clear that Davidson intends to apply the criterion in all counterfactual situations. He does, however, consider at least one counterfactual situation in discussing it.



which it is not assumed that every prior event in the history of the world that is causally linked with it is the same.<sup>7</sup>

Suppose I get my thoughts about aluminum from reading books. In the counterfactual situation I see identical-looking print. The salient difference lies at the end of a long causal chain: the counterfactual chain goes back to some other metal; the actual one goes back to aluminum. On the view we are considering, every event in the two chains must be different, a different particular. So the physiologically characterized events in my brain caused by reading the print cannot be token-identical in the two cases. What is objectionable about this view is that it makes the individuation of brain events depend on matters that are irrelevant to the physiology of the brain. I know of nothing in our explanatory practices that would support such a metaphysical view.

As far as I can see, in numerous cases it is possible for all physical (token-) events that are candidates for identification with intentional mental events to remain the same in counterfactual situations while the relevant mental events, the thoughts, differ. I do not see the slightest plausibility in the idea that there are always physical events—identifiable with the mental events and specifiable in the natural sciences—whose identity will, under all possible counterfactual circumstances, vary exactly when the mental events vary. At any rate, objections of the sort Davidson raises need to show, in the light of the thought experiments, what differences in physical events might plausibly be token-identical with the different mental events.

There is another element in Davidson's resistance to my argument. He holds that an appreciation of the external factors that enter into our common ways of identifying mental states does not discredit a token-identity theory. He cites sunburn as an example of a state whose identification is environmentally dependent. Davidson thinks, plausibly, that a particular sunburn is token-identical with a

<sup>7</sup> Davidson, ('Knowing one's own Mind', 451–453) concedes that events in the individual's body can be type-identical in the relevant actual and counterfactual situations. But he gives no reason for thinking that events in the individual's body cannot be token-identical—the same individual events—between actual and counterfactual situations.

There is an unexplained suggestion, p. 453, that he thinks that 'essentialist' assumptions might be playing a role in my argument. I do not mind being committed to some types of essentialism. But my argument does not depend on essentialism. All modal claims (including the one in the second premise) can be interpreted as supported by the best available methods of individuation. In the first premise I have not made a claim about essence or necessity at all, only one about possibility. To counter this claim, Davidson would have to appeal to a certain impossibility. He would have to deny that it is possible that an event could be the same token event if, counterfactually, its causal ancestry differs in any of the ways needed to yield different thoughts. This denial would itself seem to me to suggest an extreme form of essentialism about individual events. It suggests that the entire causal ancestry of an event is always part of the essence of that event. The denial could be reparsed in terms of methods of individuation (cf. above), but it seems to me to have no support in our actual explanatory practices. Davidson does think that events are the same if and only if they have the same causes and effects (cf. note 8). But for the present purposes, he needs to hold that it is impossible for an event to have been the same if it had had any differences in causal history. As far as I know, he has not defended this modal thesis. I suspect that it is indefensible.

particular state of the skin. And he thinks that mental states and events are broadly analogous to sunburn.

I think the sunburn example irrelevant to my argument. My argument does not claim that no non-individualistically characterized states or events can be identified with states or events in the individual's body. It claims only that some mental states and events cannot be.<sup>8</sup> The state of being sunburned is, I agree, token-identical with a physiologically specifiable state (instance) of the skin. But the state of being sunburned is relevantly unlike a state of belief, or a thought-event. In the first place, we know how to identify sunburned states of the skin in systematic and explanatory ways that are completely independent of the fact that they are sunburns. That is part of what makes plausible the identification of a sunburn with a physiological state of the skin. We can see the effect of the sun on the skin, and give a chemical or physiological account of that effect. Significantly, this account need not make any assumptions about whether a celestial body caused that effect. We have no such ways of identifying states of the body that (putatively) are beliefs, independently of assumptions about the beliefs. I think that this difference reflects our ignorance of the relation between beliefs and states of the body. It is not incompatible with an identity theory, but it renders questionable the easy analogy to cases where identities are obvious.

In the second place, the analog of my argument would not work for sunburns in the way the argument works for thoughts. The analogue of the first premise seems true for sunburn. That is, the very same (instance-identical) state of the same skin, specifiable in physiological terms, could (metaphysically could) have been caused by something other than the rays of the sun. This seems true because there are uncontroversial and obvious ways of identifying sunburns with physiologically specifiable states (state-instances) of the skin. These specifications are deeply explanatory and seem to be individually relevant. Yet they do not presume anything about those states deriving from the sun's rays.<sup>9</sup>

<sup>8</sup> Ibid. 452 says I may make the mistake of thinking that the thought experiments by themselves are incompatible with the token-identity theory. His further discussion seems to suggest that I do make this mistake. I have never thought this, and I have never written anything that entails it. In a later paper, "The Myth of the Subjective", in M. Krausz (ed.), *Relativism: Interpretation and Confrontation* (Notre Dame, Ind.: University of Notre Dame Press, 1989), 159–172, Davidson says that the argument behind my denial of the token-identity theory 'assumes that if a state or event is identified (perhaps necessarily, if it is a mental state or event) by reference to things outside of the body, then the state or event itself must be outside the body, or at least not identical with any event in the body'. I do think that certain mental events that are necessarily individuated by reference to things outside the body are not identical with any event in the body describable via kind terms of the physical sciences. But I have never taken this view as an assumption. I have argued for it. Davidson's discussion does not go to the heart of our disagreement. What follows should make the point clearer.

<sup>9</sup> In my view, therefore, the sunburn example does not line up very well with Davidson's criticism of my argument—since he wishes to deny premise (1). Davidson very likely denies that the analog of (1) holds for sunburn. Cf. Davidson, "Knowing One's Own Mind", 451–455, where he concedes only that the state of the skin might remain type-identical. But he gives no reason for holding that an instance-identical state of the skin could not have been caused by anything other than the sun. Perhaps he thinks that states are analogous to events in (necessarily) being instance-identical only if

On the other hand, the analog of premise (2) seems to fail for sunburn. I think that we have no grounds to think that if a physiologically specifiable state (instance) of the skin had not been a sunburn, it would have been impossible for it to have been the same state (instance) as one that is. Again, our ability to identify states of the skin with sunburns and fit them into a systematic, individually relevant scheme of explanation that prescind from anything about the sun supports this view. The kind *sunburn*, by contrast, fits into a much less systematic, informative, and important explanatory scheme—if it fits into any such scheme at all. Its claim to individuate state-instances across counterfactual situations is far less strong than the physiological scheme. Similarly, its claim to provide a fundamental form of individuation (as premise (2) claims for intentional contents) is substantially less strong than the claim of mental kinds. The failure of analogy, in these respects, between sunburn and thoughts makes the example incapable of undermining my argument.

Whereas Davidson doubts premise (1) of the argument, some token-identity theories have been taken as denying premise (2). It is common to hold that intentional states and events can be type-identified with functional states and that these states could be realized by a variety of physical states or events. The realizing physical states and events are supposed to be instance- or token-identical with the mental states or events. This view does not entail a denial of (2). But it is usually also assumed, at least implicitly, that the identity of the underlying physical states and events is independent of their exact functional role in a given system. If this is assumed, then this sort of theory would be committed to denying (2).

Such a theory seems very unpromising. There is little in our explanatory practice to encourage identification of intentional mental types with non-intentionally specified functional types.<sup>10</sup> The identification of intentional mental tokens with tokens specified in the natural sciences has, I think, even less clear support in explanatory practice. But what I want to centre on is the main reason for accepting (2).

The system of intentional content attribution is the fundamental means of identifying intentional mental states and events in psychological explanation and in our self-attributions. In fact, we have no other systematic way of identifying such states and events. Davidson holds that (2) is intuitively obvious and denials of it are absurd. I agree. But I think that it can also be justified by noting that it is our only way of individuating intentional mental events that provides systematic

they have the same causes (cf. note 6). But I see no metaphysical or even physiological impossibility in thinking of a given state-instance as having been arranged by light rays with exactly the same physical powers and vectors as the sun's which, however, did not emanate from the sun. Whether Davidson holds the analogue of premise (2) for sunburn is unclear to me. It appears, however, that he probably would hold that no state-instance of a sunburn could be caused by something other than the sun and be the very same state-instance. As noted below, I find this implausible.

<sup>10</sup> For a recent discussion of reasons for this charge, cf. Hilary Putnam, *Representation and Reality* (Cambridge, Mass.: MIT Press, 1988).

understanding, description, and explanation of mental events and intentional activity. Such cognitive practices are our most reliable way of knowing the nature of what exists.

I know of no plausible or even serious arguments against (2).<sup>11</sup> But failure to hold it firmly in mind is a major source of the misguided worries about epiphenomenalism. Insofar as one allows oneself to think that intentional contents are only contingently associated with a thought-event, one is in a position to imagine that one can hold constant a neural event supposedly identical with a mental event, while imagining the mental/physical event to have a different content. Supposing that the causal laws connecting the neural processes are sufficient for their effects, one is led to wonder wherein the intentional properties are causally relevant. This conceit can naturally be seen to lie behind the view that the intentional content of mental states and events is a mere epiphenomenal ‘aspect’ of those events—the view that I noted in the first paragraph of this paper.

But in fact, such content is the explanatory and identificatory centre of those events. We have little else to go on in talking about the causal powers and ontology of the mental. Systematic, informative, important explanatory schemes—like our mentalistic one—usually (there are special cases) make the strongest claim for providing individuating descriptions that indicate what is essential to the identities of individuals, particulars, or instances.<sup>12</sup> It seems to me that any metaphysical theory that seeks to illumine mental causation or the ontology of mental events that denies (2) is hopelessly misdirected.

As I have been noting, systematic, informative, important explanatory schemes of events and states are also our strongest indications of causal relevance. The idea that the causal relations described in psychology are really or most fundamentally causal only under some other description seems extremely tenuous and doubtful. I see no reason to think that there is anything in the idea, now common among philosophers, that in some sense the ‘real’ causal work is being done at a lower level. I also see no reason to think that we can understand mentalistic causation through some analysis of supervenience (although I think that understanding the sense in which mental events supervene on the physical is an important enterprise). Our understanding of mental causation derives not primarily from re-descriptions in physical terms.<sup>13</sup> It derives

<sup>11</sup> I have heard it suggested that although (2) is not true, it is true that there could not be two events that are mentally the same if they had different contents. (2) is still denied because the same physical event could allegedly be identical with a mental event that had one content, in the actual situation, or with a mental event that had another content, in the counterfactual situation. It is not the same thought content, but it is the same thought/neural event. This subtlety does not seem to me to count for much in defending opposition to (2). The remarks in the preceding paragraph still seem to apply.

<sup>12</sup> Again, one can reparse talk about essence into talk about the most fundamental method for identifying individuals and reidentifying them under counterfactual considerations; cf. note 7.

<sup>13</sup> Or worse yet, in syntactical terms. One might well grant that many psychological states and events involve operations on syntactic entities. One may also see such inner mental syntactic items

primarily from our understanding of mentalistic explanation. This understanding is largely independent of reference to the underlying processes.

Davidson's profoundly stimulating argument for a token-identity theory seems to me flawed by a subtle analog of the tendency to underrate the centrality of cognitive practice in understanding mental causation and ontology.<sup>c</sup> The argument partly relies on the premise that causal relations between events must be backed by (entail the existence of) laws of a complete, closed system of explanation, where the predicates of these laws are true of the events that are causally related. Davidson conjoins this premise with the assumption that there are causal relations between mental and physical events and the assumption (for which he argues) that mentalistic explanations cannot themselves be part of a complete, closed system of explanation. From these three premises he concludes that mental event-tokens must fall under predicates of physical laws that do form part of a complete, closed explanatory system. As far as I can see, these must be the laws of physics.

I think that we do not know, and cannot know apriori, that causal statements entail the existence of laws or explanatory systems that have such specific properties. We cannot know apriori this much about the form of the laws of nature, described by any science. Nor can we know apriori the relations among the ontologies and causal schemes of the various sciences. That is, we cannot know apriori that the laws, or nomological generalizations, of psychology describe events that have any very specific relation to events that fall under the laws of any other science. I think that there is no reason, much less any apriori reason, to accept Davidson's first premise. Thus we cannot know apriori that mental events that are causes fall under any other (non-mentalistic) sort of law.

Our best guide to the nature of mental causation and ontology lies in understanding our best means of explaining and describing mental events. Our actual cognitive practice lends little credence to the idea that intentional mental events fall under exceptionless physical laws, or instantiate physical descriptions of the most fundamental natural science.

Let us return to the argument against the form of token-identity theory that I described at the outset (a form that Davidson's argument was intended to support). Since premises (1) and (2) seem strongly plausible and free of serious counter-arguments, the argument seems to me forceful and sound. The most

as expressing intentional content. But the idea that the brain is sensitive only to syntactic shape, not to content, seems to me a misleading metaphor. The brain is sensitive to neural forms. One may think of there being causal relations among syntactic entities, but this is already to describe causation at a higher level of abstraction than that of brain physiology. If causation can be described at that level, it can also be described at the level of intentional content. I see little ground for seeing the syntactic properties as prior to intentional properties in the order of causal relevance. Syntax is attributed to serve explanations that are fundamentally intentional. In my view, there is no non-artificial sense in which syntax is ontologically or causally more secure.

<sup>c</sup> D. Davidson, "Mental Events", in L. Foster and J. Swanson (eds.), *Experience and Theory* (Amherst, Mass.: University of Massachusetts Press, 1970), 79–101; repr. in Davidson, *Essays on Actions and Events* (Oxford: Clarendon Press, 1980).

common metaphysical ground for worries about epiphenomenalism seems on weak ground.

There are less committal forms of materialism about mental events. All such forms must take mental event-tokens to be physical event-tokens that are not instances of kinds described by any of the natural sciences. Perhaps this is not an intrinsically implausible claim. Particular wars, avalanches, thunderstorms, meal-cookings may not fall under any natural event-kind describable in any natural science. I doubt that there is ever any one definite token event that instantiates kinds of a natural science, even physics, with which an event-token of these kinds can be identified. Yet they are clearly physical events. Maybe mental events are like that.

On the other hand, these analogies seem merely hopeful. The relevant descriptions are uncontroversially physical. The problem with identifying them with events described in physics lies merely in their complexity and in the fact that they fall into patterns that are salient for macro-descriptions, but not particularly useful for systematic explanation. By contrast, psychological states do fall into systematic patterns which explanation can make use of. But these patterns do not seem to involve ordinary physical properties (like mass, energy, composition, and so on) that physical explanations and descriptions make use of.<sup>14</sup> So the view that these are, in the ordinary sense, physical patterns seems doubtful.

In any case, such forms of materialism provide little new insight into mental causation, as long as the relevant physical events with which the mental events are supposedly token-identical remain unidentified in non-mentalistic terms. There is little in current cognitive practice to encourage the view that any such descriptions of central intentional state-instances or event-tokens are forthcoming. So even if such a liberalized version of materialism were true, it would offer little help in understanding mental causation.

3

The problem about mental causation that is usually raised as a means of motivating an appeal to materialism is that of explaining a mechanism that would make possible causal interaction between two such different things as a physical event (or substance) and a mental event (or substance). Materialists hold that the problem disappears if mental events are seen as physical. This problem was posed by Descartes; and he professed himself baffled by it. The problem was acute for Descartes because he viewed mental and physical entities as substances in

<sup>14</sup> The composition relation is probably the critical one. There are forms of materialism that maintain that all objects are decomposable into inorganic physical particles. Apart from the fact that these forms do not apply very well to properties or events (not to speak of objects like numbers, intentional contents, methods), they make a claim for the relevance of physical composition to our understanding of mental entities that seems to me (so far) quite unsupported by anything that we know.

the old-fashioned sense—entities that were not dependent on anything else for their nature or existence. But mental events and other mental entities like minds are not substances in the old-fashioned sense. So there is for us no antecedent problem of admitting various sorts of constitutive and existential dependency of the mental on the physical. It is not obvious, however, that this dependency need involve material constitution—the mental events being identical with or made up of physical events.

I have no satisfying response to the problem of explaining a mechanism. But I am sure that there is less reason to think it a decisive consideration in favor of materialism than is often thought. What is unclear is whether the question is an appropriate one in the first place. Demanding that there be an account of mechanism in mind–body causation is tantamount to demanding a physical model for understanding such causation. It is far from obvious that such a model is appropriate. It is not even obvious why any model is needed. The argument I have just cited presents no clearly formulated problem about mental causation that need force us to embrace materialism, including the computer model's version of materialism, as a solution. The demand for a mechanism is tantamount to an implicit demand for a materialist solution.

A better articulated argument for materialism along similar lines goes as follows. Physical effects are caused by prior physical states or events according to approximately deterministic physical laws. Mental causes bring about physical movements of our bodies. If such causation did not consist in physical processes, there would be departures from the approximately deterministic physical patterns described by physical laws. It would interfere with, alter, or otherwise 'make a difference' in the physical outcomes. But there is no reason to think that this happens. Physical antecedent states suffice for the physical effects. Appeal to mental causation that does not consist in physical causation appears, on this reasoning, to require us to doubt the adequacy of current forms of physical explanation, even within the physical domain. So such appeal should be rejected.

This reasoning seems to me to have some force. But I think that it is not as forceful as it may appear. Why should mental causes alter or interfere with the physical system if they do not materially consist in physical processes? Thinking that they must, surely depends on thinking of mental causes on a physical model—as providing an extra 'bump' on the effect. The idea seems to be that a cause must transfer a bit of energy or exert a force on the effect. On such a model, mental causes would deflect a physical effect off the course a physical cause alone would set it on. Cases where the mental cause and the physical cause yield the same physical outcome—cases of 'overdetermination'—will be seen, on such a model, as coincidences that must be abnormal. So appeals to overdetermination by mental causes and physical causes will seem unattractive as a gloss on the general failure of mental causes to interfere with or alter the physical outcomes expected on purely physical grounds.

But whether the physical model of mental causation is appropriate is, again, part of what is at issue. As we have seen, one can specify various ways in

which mental causes ‘make a difference’ which do not conflict with physical explanations. The differences they make are specified by psychological causal explanations, and by counterfactuals associated with these explanations. Such ‘differences’ made by psychological causes do not require that gaps be left in physical chains of causation. They do not seem to depend on any specific assumptions at all about the physical events underlying the mental causes.

I think that we have reason, just from considering explanatory goals and practice—before ontology is even considered—to think that mentalistic and physicalistic accounts of causal processes will not interfere with one another. Part of the point of referring to mental states lies in explaining intentional activity that involves (or is identical with) physical movement. A man’s running to the store is explained by his believing that his child would suffer without the needed medicine and by his decision not to wait on a doctor. We think that the man’s running is caused by the formation of his belief and by his decision.

It would be perverse to think that such mental events must interfere with or alter, or fill some gap in, the chain of physiological events leading up to and including the movements of his muscles in running. It would be perverse to think that the mentalistic explanation excludes or interferes with non-intentional explanations of the physical movement. I think that these ideas seem perverse not because we know that the mental events are material. They seem perverse because we know that the two causal explanations are explaining the same physical effect as the outcome of two very different patterns of events. The explanations of these patterns answer two very different types of inquiry. Neither type of explanation makes essential, specific assumptions about the other. So the relation between the entities appealed to in the different explanations cannot be read off the causal implications of either or both types of explanation. The perversity of thinking that mental causes must fill gaps in physical chains of events probably has its source in traditional dualism, or in libertarian worries about free will. But the perversity remains regardless of its source.

The upshot of this reasoning is that we have no ground for assuming that the failure of mental causes to interfere in the physical chain of events must be explained in terms of mental causes’ consisting in physical events. Interference would be surprising, given antecedent assumptions about mental and physical explanation. So non-interference is in no need of explanation in ontological terms.

There are surely some systematic, even necessary, relations between mental events and underlying physical processes. We have good reason to believe that mental processes depend on underlying physical processes. By probing or damaging parts of the brain we can bring about, affect, or ruin mental states and processes. But the relations of identity and physical composition are relations that have specific scientific uses. For example, we explain the behavior of a molecule in terms of the behavior of its component parts. It is far from clear that these compositional relations have a systematic scientific use in bridging psychology and neurophysiology (cf. note 15). They are guesses about what



sorts of relation might obtain. But they seem to me just one set of possibilities for accounting for relations between entities referred to in these very different explanatory enterprises. What form an inter-level account might take seems to me to be an open question.

I find it plausible to believe in some sort of broad supervenience thesis: no changes in mental states without some sort of change in physical states. But the inference to materialism is, I think, a metaphysical speculation which has come, misleadingly, to seem a relatively obvious scientific or commonsensical bromide.

As long as mentalistic explanation yields knowledge and understanding, and as long as that explanation is (sometimes) causal, we can firmly believe that mind–body causation is a part of the world. The primary way of understanding such causation is by understanding mentalistic causal and explanatory statements in the ordinary, non-philosophical sense of ‘understanding’. How much more illumination philosophy or neuro-psychology can offer remains to be seen. At any rate, mentalistic explanation and mental causation do not need validation from materialist metaphysics.

It seems to me that philosophers should be more relaxed about whether or not some form of materialism is true. I think it a thoroughly open—and not very momentous—question whether there is any point in insisting that mental events are, in any clear sense, physical. Maybe science will never make use of anything more than limited correlations with the lower, more automatic parts of the cognitive system. Maybe identities or part–whole relations will never have systematic use. Maybe the traditional idea of a category difference will maintain a presence in scientific practice. What matters is that our mentalistic explanations work and that they do not conflict with our physicalistic explanations. As philosophers, we want a well-founded understanding of how these explanations, and their subject matters, relate to one another. But it serves no purpose to over-dramatize the conflict between different ontological approaches or the merits of the materialist approach.

The flood of projects over the last two decades that attempt to fit mental causation or mental ontology into a ‘naturalistic picture of the world’ strikes me as having more in common with political or religious ideology than with a philosophy that maintains perspective on the difference between what is known and what is speculated. Materialism is not established, or even clearly supported, by science. Metaphysics should venture beyond science with an acute sense of its liabilities.

I have argued that epiphenomenalism need not be seen as a serious metaphysical option even if materialism is true. The probity of mentalistic causal explanation is deeper than the metaphysical considerations that call it into question. I have

also argued that materialist metaphysics is not the most plausible starting point for reasoning about mind–body causation. Explanatory practice is.

I think it more natural and fruitful to begin by assuming, defeasibly perhaps but firmly, that attributions of intentional mental events are central to psychological explanation both in ordinary life and in various parts of psychology. We may also assume that intentional mental events are often causes and that psychological explanation is often a form of causal explanation. Given these assumptions, the ‘worry’ about epiphenomenalism seems very remote. For if intentional mental events are type-individuated in terms of their intentional ‘aspects’—if those aspects are the fundamental explanatory aspects—and if such events enter into causal relations and are cited (in terms of those aspects) in explanations, then there seems to be every reason to conclude that those aspects are causally efficacious. None of the metaphysical considerations advanced in current discussion seem to me remotely strong enough to threaten this conclusion.<sup>15</sup>

I shall conclude by mentioning a further reason for thinking that intentional events are, as such, causally efficacious. Why is it that philosophers who discuss epiphenomenalism are worried about it? Why should one not take a more disinterested attitude?

Part of the answer lies in the fact that many philosophers instinctively feel that if epiphenomenalism were a consequence of their metaphysics, their metaphysics would be in trouble.<sup>16</sup> This is good philosophical sense. Outside our philosophical studies, we all know that epiphenomenalism is not true.

But there is a deeper reason why epiphenomenalism should seem intellectually unattractive. Much of the interest of psychological explanation, both in psychology and in ordinary discourse, lies in helping us understand ourselves as agents. Causal implications are built into our intentional concepts and intentional modes of explanation. We think that we make things happen because we make decisions or will to do things. We think that we make assertions, form theories, and create cultures, because we think certain thoughts and have certain goals—and we express and fulfill them. In this context, we identify ourselves primarily in terms of our intentional mental aspects—our wants, our thoughts, our values. Our agency consists in our wants’, willings’, thoughts’, values’ as such (under these ‘aspects’) having some sort of efficacy in the world. Our mental events’ having the intentional characters that they have is, in individual instances, what we define our agency in terms of.

<sup>15</sup> The fact that intentional mental events are appealed to, as such, in explanations by itself differentiates them from phenotypes. Phenotypes are explananda in biology, but intentional mental events are part of causal explanations in psychology and everyday life—not simply the explananda for explanations. It is a mark of much of the discussion in this area that such simple differences in explanatory status are not so much as mentioned. The metaphysics of the situation is often discussed in virtual isolation from explanatory practice.

<sup>16</sup> Cf. N. Block, “Can the Mind Change the World?”, in G. Boolos (ed.), *Meaning and Method: Essays in Honor of Hilary Putnam* (Cambridge: Cambridge University Press, 1990), 137–170, for a clear statement of this view.

Most of our intellectual and practical norms and evaluations presuppose that we are agents. If our willing or deciding made something happen, but that event's being a willing or deciding were not causally efficacious (so that the efficacy resided in some underlying neural property), then the agency would not be ours. If our theoretical deliberations were not ours to control, we could not see ourselves as being the authors of our theories; nor could we criticize ourselves as deliberators. Most normative evaluations of our intellectual and practical activities would be empty.

If intentional psychological explanation 'made sense' of what we did, 'rationalized' it, but did not provide insight into the nature of any causal efficacy, it would lose much of its point. It would provide no insight into the various forms of agency that give life its meaning and purpose, and psychology its special interest.

I am not here asserting that it is inconceivable that psychological explanation could break down—or at least be very much more limited than we conceive it as being. I think that the question of conceivability is quite subtle and complex. The point is that this form of psychological explanation has not in the least broken down. It works very well, within familiar limits, in ordinary life; it is used extensively in psychology and the social sciences; and it is needed in understanding physical science, indeed any sort of rational enterprise. We have reason to believe that it provides explanatory insights. But then, as I have noted, there is a deep connection between intentional psychological explanation and the view it provides of ourselves as agents, on one hand, and the attribution of some sort of causal efficacy to intentional mental states and events as such, on the other.

I think that these are not the only reasons for this form of attribution. But even if they were, it would leave the attribution so deeply entrenched that there is no real hope that epiphenomenalism could become a credible view. Epiphenomenalism is better seen as an instrument, like scepticism, for clarifying our most deeply held beliefs. It seems to me, however, that the traditional scepticisms about agency, will, and responsibility are more penetrating tools for this purpose.

## **Postscript to "Mind–Body Causation and Explanatory Practice"**

"Mind–Body Causation and Explanatory Practice" is out of step with the philosophical temper of the times. I take a distant, sceptical attitude toward the prevailing physicalism, or materialism, in philosophy of mind. I also defend my old argument, first stated in "Individualism and the Mental" (Chapter 5 above), that one popular form of materialism, the token identity theory, is false.

I attempt no satisfying account of the mind–body problem. My primary interest lies in articulating dissatisfaction with the particular approach to the problem that has dominated discussions since the mid-twentieth century.

I believe that discussions of the mind–body problem often show poor perspective on what we know and what we do not know. Metaphysical positions—like materialism or physicalism—are held in common among many philosophers with a firmness and fervor that are out of proportion to the strength of the grounds for holding them. Such positions are not clearly supported by explanations in common sense or the sciences. Nor do they have strong intuitive support.

There is, of course, a long-standing, justified, general sense that the mind depends on the body. However, materialist views in philosophy make very particular claims about the relations between psychological events or properties and physical events or properties. Such views require relations of identity, constitution, realization, or the like. No one of these specific relations is clearly supported by actual scientific explanation. I think that all lack strong intuitive support—beyond the support for the generalized sense of dependence.

In the article, I held that the issues over mental causation that had begun to arise from within materialist points of view showed the metaphysical situation to be less clear-cut and more difficult than common philosophical opinion took it to be. I maintained that metaphysical positions in this area should be occupied and reflected upon with considerably more diffidence, and with a greater openness to reflecting on alternatives.

Similarly, the metaphysical view epiphenomenalism, which is flatly incompatible with what we know from scientific and commonsense explanations, is widely taken as a genuine contender for the truth, although, unlike physicalism, it is not widely held. The reason why the view is taken seriously is that it can appear to be implied by some of the metaphysical opinions that are themselves too firmly held. Here again, I maintained that a great deal of philosophy had lost perspective on the distinction between what we know and understand,

on one hand, and what we are in the position of groping to understand, on the other.<sup>1</sup>

I believe that these points remain valid. My repeating them here is witness to their having had rather little effect. I believe that part of the problem is that large shifts in philosophical attitude usually proceed only slowly. In some cases, however, misunderstanding has helped block adequate appreciation of the points that I made. I want to discuss a response that embodies both a fair amount of misunderstanding and a version of the syndrome that provoked my scepticism.

Writing as if my criticism of materialist metaphysics were a criticism of doing metaphysics at all, Jaegwon Kim responds to my paper by arguing that there is indeed a problem of mental causation and that a metaphysics that engages it is worth doing. Those points are true. Kim's spirited claims in this vein are irrelevant to what I wrote.

Kim quotes my statement,

Materialist metaphysics has been given more weight than it deserves. Reflection on explanatory practice has been given too little.

He associates this statement with the much stronger claim of Lynne Rudder Baker:

If we reverse the priority of explanation and causation that is favored by the metaphysician, the problem of mental causation just melts away.<sup>2</sup>

Kim asks:

Would the problem of mental causation take its leave if we did less metaphysics, as Burge and Baker urge, and instead focused our attention on psychological explanation?<sup>3</sup>

This take on my view is very much mistaken. I do not accept Baker's claim that the problem of mental causation would 'melt away' if one shifted perspective. I did not, and do not, think that any such problem can be made to 'take its leave' by focusing on psychological explanation, and doing less metaphysics. I do not advocate doing less metaphysics.<sup>4</sup> I have no objection to counting the problem of understanding mental causation as a *metaphysical*

<sup>1</sup> I applied these methodological points to particular issues regarding mind–body causation. I believe that parallel points apply even more strongly to certain other areas of metaphysics. For the most part, philosophy of mind has taken at least rough account of what (little) is known about the mind and its relation to the brain or body. In some other areas of metaphysics, the discussion has proceeded without any serious contact with what science has had to say on the relevant topic. I have in mind, for example, discussions of the nature of time, the relation between existence and time, the nature of physical bodies, implications for contact in body–body causation, and much of the discussion of causation itself.

<sup>2</sup> Lynne Rudder Baker, 'Metaphysics and Mental Causation', in J. Heil and A. Mele (eds.), *Mental Causation* (Oxford: Clarendon Press, 1993), 92–93.

<sup>3</sup> Jaegwon Kim, *Mind in a Physical World* (Cambridge, Mass.: MIT Press, 1999), 60–62.

<sup>4</sup> The scope of Kim's 'as Baker and Burge urge' is ambiguous. His assumption about what I urge is mistaken on either interpretation.

problem. I believe that very few philosophical problems are easily solved or deflated. I do not believe that problems regarding mental causation are among these few.

I recognize that there are difficulties in achieving a satisfactory understanding of mental causation, though I think that the ways these difficulties have usually been posed, from Descartes onward, have been unproductive. I think that the problem has usually been posed so as to suggest that there is some definite conflict or tension in the very notion of mind–body causation. I believe that this suggestion has never been made good. It was not made good by Descartes or by Princess Elisabeth of Bohemia (who raised the issue with Descartes). It has not been made good by modern materialists. Beginning one’s inquiry by taking such a suggestion seriously, rather than beginning with a firm grip on what we know from good psychological and physical causal explanations, has tended to distort philosophical discussion. My paper was partly a complaint about such distortion. I believe that by reflecting more on explanation in science and common sense, one can gain a better starting point and perspective on the problem of mind–body causation.

As noted, I also maintained that a particular view, epiphenomenalism, has been taken more seriously, as a contender for the truth, than it deserves. I held that epiphenomenalism is a non-starter. I argued that epiphenomenalism can serve at best as a foil in trying to better understand mental causation. One can use any valid argument that leads to epiphenomenalism as a *reductio* of the argument’s set of premises.

Kim writes that he agrees with my attitude toward epiphenomenalism. He maintains, ‘the problem of mental causation is ... the problem of showing *how* mental causation is possible, not *whether* it is possible’. He continues:

The issue is *how to make our metaphysics consistent with mental causation*, and the choice that we need to make is between various *metaphysical alternatives*, not between some recondite metaphysical principle on the one hand and some cherished epistemological practice or principle on the other. This of course is not to say that metaphysics and epistemology are necessarily independent ...<sup>5</sup>

As we shall see, however, Kim does not disengage his own reasoning from talk of mental causation’s being in ‘jeopardy’—jeopardy engendered by what I believe to be insubstantial metaphysical considerations. In fact, I believe that because of specific aspects of Kim’s eventual reductionism, his own position amounts to a form of epiphenomenalism.

Mental causation does need to be better understood, particularly in relation to physical causation. I conceive this problem in terms of better understanding rather than in terms of explaining the possibility of mental causation. I think its possibility is best explained by reflecting on its actuality rather than by appealing to putative metaphysical principles, as Kim does. Perhaps this difference

<sup>5</sup> Kim, *Mind in a Physical World*, 62.

is merely one of strategy. I do not deny that metaphysical exploration and speculation can be fruitful when it is done with a keen sense of its limitations.

I have more substantive differences with the last part of Kim's statement. Part of my point was that we cannot assume that we need to *make a choice* between various metaphysical alternatives if the available alternatives do not have a firm rational or evidential basis. Metaphysics should sometimes be carried on in a more exploratory and speculative spirit.

The issue is not between metaphysics and epistemology, two branches of philosophy. It is not a matter of making a choice between doing metaphysics and doing science, as the quoted paragraph seems to suggest my view to be. The issue is that certain forms of metaphysics do not keep in perspective what we know and what we do not know—including what we know from science. Such forms rely on metaphysical principles that are not rationally or empirically supported. Most of what we know (or are warranted in believing) about mental causation resides in causal explanations in science and common sense. In my view, metaphysics in this area has progressed very little beyond them.

Kim claims that our understanding of belief–desire explanation as causal derives from Davidson's argument that reasons are causes:

If, as Burge says, we 'may assume' that belief–desire explanation is a form of causal explanation, we owe this license substantially to Davidson. What carried the day for the causal view was Davidson's philosophical argument, not the pervasiveness of our explanatory practice of rationalizing actions in terms of belief and desire. There was no disagreement on the explanatory practice; the debate was about its nature and rationale.

Kim seems to think that explanatory practice in psychology and common sense was innocent of evident causal commitments until Davidson's philosophical (indeed Kim holds, metaphysical) arguments came along to indicate their causal character.<sup>6</sup>

I admire Davidson's defense of regarding psychological explanation as causal.<sup>7</sup> I think, however, that Kim's account takes too narrow a view of the dialectical situation in which Davidson's defense was mounted. Davidson resisted a set of aberrant philosophical views that systematically either ignored or were suspicious of scientific explanation in psychology and that offered what was in fact a revisionist view of commonsense explanation. Ordinary and scientific explanatory practice did not need Davidson to show it to be causal. Philosophy, during one of its less admirable periods, did.

Kim seems to think that actual scientific explanation in psychology is *prima facie* neutral as to whether psychological events or properties are causally efficacious, and that such explanation needs metaphysics to determine the issue. I

<sup>6</sup> Kim, *Mind in a Physical World*, 62–63. The quote is from p. 63.

<sup>7</sup> Donald Davidson, 'Actions, Reasons, and Causes', *The Journal of Philosophy*, 60 (1963); repr. in his *Essays on Actions and Events* (Oxford: Clarendon Press, 1980).

think that such a view is completely mistaken. We are entitled to assume that psychological explanation is causal unless powerful arguments arise to the contrary. We do not need metaphysics to license the assumption. Mental causation is solidly supported by science (not to speak of common sense). No metaphysics—at least, no metaphysics that has emerged so far—is in a position to put mental causation ‘in jeopardy’.

I think that Kim’s position is yet another example of greater confidence in a metaphysics about the mind–body problem than the relevant arguments warrant. I would like to develop this view in more detail.

Kim thinks that the principal problem for understanding mental causation is what he calls ‘the exclusion problem’. The problem is: ‘Given that every physical event that has a cause has a physical cause, how is a mental cause also possible?’<sup>8</sup>

As noted, I think that the first answer to this question lies in looking at actual mental causation—at the specifics of causal explanations in psychology and common sense. We need to recognize from the outset that the psychological properties appealed to in causal psychological explanations of physical events are among the causal factors. No ground to doubt the legitimacy of appeal to them as causes has so far been raised. Given that we know that there is neuro-physical causation of the same movement, we know that the psychological and the neuro-physical causation do not compete. I believe that we have no intuition that we need take seriously that they do compete. I think that intuitions that suggest such competition derive from questionable, poorly supported metaphysics or from dubious metaphors. Any such intuitions are in conflict with what we know from scientific and commonsense explanations.

Kim wants to press a puzzlement or a sense of ‘tension’ here that I think is not well motivated. He thinks that it is natural to regard the presence of the physical cause as a *threat to exclude* the presence of a mental cause. Thus, his name for the problem. Sometimes he writes that the physical cause can make the mental cause seem ‘dispensable’ or can put it in ‘jeopardy’.<sup>9</sup>

I find this approach artificial. I think that given what we know from psychological causal explanation, we should assume from the beginning that any sense of tension is an illusion that must derive from some misunderstanding. Now in a way, Kim agrees that the sense of tension is an illusion. He wants to leverage the sense of tension into an acceptance of his reductionist view of psychological properties, and psychological explanation.<sup>10</sup> Where we differ is that I think that the sense of tension is not only an illusion, but is itself largely artificially induced.

<sup>8</sup> Ibid. 37–38.

<sup>9</sup> Ibid. 42, 45.

<sup>10</sup> Kim presents his view as a reductionist view, but the notion of reduction that he invokes seems to me not at all like reductions in science. In many respects, the view seems more eliminativist or epiphenomenalist, at least about representational states. I shall, however, continue to call it a reductionist view.



Kim states that the exclusion problem arises ‘for anyone with the kind of broadly physicalist outlook that many philosophers, including myself, find compelling or, at least, plausible and attractive’.<sup>11</sup> This is to say that the exclusion problem arises from within physicalist metaphysics. Its force or intuitiveness is, I think, dubious independently of the metaphysics. It is very doubtful that it can be used to motivate or support physicalist metaphysics, or any particular version of such metaphysics.

Kim tries to build the sense of tension between mental and physical causes into a *reductio* argument that leads to epiphenomenalism. He wants to use the argument not to support epiphenomenalism, but to support his reduction of mental properties to physical properties. I think that there is more than one dubious or very speculative step in the argument. So I think that it cannot be used to target any one premise as the mistaken one in generating epiphenomenalism.

The argument begins with a sub-argument that mental properties *supervene* on physical properties in the sense that

if something instantiates any mental property *M* at *t*, there is a physical base property *P* such that the thing has *P* at *t*, and necessarily anything with *P* at a time has *M* at that time.

Kim’s argument for this principle seems to me very strange. He bases his argument on ‘*the principle of physical causal closure*’:

If you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the nonphysical. ... If you reject this principle, you are ipso facto rejecting the in-principle completability of physics—that is, the possibility of a complete and comprehensive physical theory of all physical phenomena. For you would be saying that any complete explanatory theory of the physical domain must invoke non-physical causal agents.<sup>12</sup>

<sup>11</sup> Ibid. 30. Kim opens his book by announcing his commitment to a ‘broadly physicalist outlook’. There is no detailed discussion of what such an outlook is. Kim says only that it is an outlook according to which ‘the world’ is ‘fundamentally physical’. There are many questions to be raised about this idea and how it is supposed to apply to various cases (the mathematical ‘world’, the ‘worlds’ of value, right and wrong, beauty, rational justification, semantics, indeed mind). I find this sort of generalized rhetoric, which is certainly not peculiar to Kim, unilluminating. The rhetoric is very far from the expression of a definite, warranted belief.

<sup>12</sup> Ibid. 40. Kim cites Jerry Fodor as holding that the very intelligibility of mental causation depends on mind–body supervenience. He apparently regards the principle of physical causal closure as a way of articulating support for Fodor’s blanket and apparently otherwise unsupported claim. Cf. Jerry Fodor, *Psychosemantics* (Cambridge, Mass.: MIT Press, 1987), 42. I think that Fodor’s claim is wild, and that Kim’s way of supporting it is implausible. I think that intelligibility lies in psychological causal explanation. Such explanation may invite philosophical supplement and interpretation. But its intelligibility hardly depends on some extremely general, abstract, and scientifically idle principle like the supervenience principle. I think that scientific explanation in psychology would go on attributing mental causation intelligibly and fruitfully regardless of the truth or falsity of supervenience, or indeed regardless of the truth or falsity of physicalism. From conversation, I believe that Fodor’s present view is now closer to mine than it was when he made this statement.

Kim maintains that if mind–body supervenience fails, and if causation from the mental to the physical occurs, this principle would be false. I think that Kim’s claim here is correct, almost regardless of how the physical causal closure principle is interpreted.

The formulation of the principle does, however, leave something to be desired. What is it to be taken ‘outside the physical domain’? What is meant by ‘agents’? It appears from the context that Kim intends to include as ‘agents’ any causal factors—properties, relations, events, and so on. Normally, we would say that a theory in purely physical terms, such as a physics or a biology, would leave out psychological causal factors, including psychological events and properties. Of course, if psychological properties were themselves physical, the principle would not be violated. Many—perhaps most—materialist philosophers, however, regard psychological properties as irreducible to (and not identical with) physical properties. So the principle as Kim appears to interpret it would not be accepted even by many materialists. Kim does not argue for the principle. He simply surrounds it with misleading rhetoric.

For example, Kim misleadingly claims that the principle is a necessary condition on taking a physical theory to give a complete theory of physical phenomena. This claim is certainly untrue, given what is *ordinarily* meant by the idea that one can give a complete physical theory of physical phenomena. What is ordinarily meant is that physical theory can explain physical phenomena by reference to causes specified in physical vocabulary, and by reference to laws expressed in physical vocabulary, where there are no gaps in the causal chains or causal explanations. This normal understanding of the completeability of the physical sciences does not prejudge any relation that those sciences bear to the human or psychological sciences, except that the latter sciences will not specify causes that intrude on the course of physical causation or on the lawfulness of the physical laws. (This understanding of the completeability of the physical sciences does not even entail supervenience.) Kim’s principle makes unnecessary his complex argument for reduction of the psychological to the physical. Assuming that mind–body causation is not epiphenomenal, the principle already entails reduction or elimination of the psychological. I regard Kim’s argument for supervenience as implausible because his premise is so strong.<sup>13</sup>

Should we accept supervenience? I find it plausible. Yet I know of no interesting argument for it. I do not think it irrational to suspend belief about it,

<sup>13</sup> In a subsequent book Kim states a closure principle that is similar to what I just (independently) wrote is the ordinary understanding of what is meant by the completeability of the physical sciences. Kim does not state that he has revised his understanding of closure. I think that he must have done so, however. For the new principle is not taken to support supervenience. In the later book Kim accepts supervenience as a separate basic principle, rather than something to be argued for from any closure principle. Cf. Jaegwon Kim, *Physicalism, Or Something Near Enough* (Princeton: Princeton University Press, 2005), 15–16, 21–22, 43–44. This procedure seems to me an improvement on the earlier work. The improvement does not affect anything except for my criticism of the earlier closure principle.

given the extreme generality of its modal claim, given that there is no strong argument for it, and given that it is neither evident in itself nor strongly supported by scientific evidence. I will accept it for the sake of argument. In what follows, let us remember that its epistemic credentials are not very strong.

Kim gives an argument that assumes supervenience and purports to lead to epiphenomenalism. Suppose that an instance of mental property  $M$  causes another mental property  $M^*$  to be instantiated. Given supervenience,  $M^*$  has a physical supervenience base  $P^*$ . At this point, Kim attempts to get the reader to share his sense that there is a tension between mental causation of  $M^*$  and the relation between the supervenience base  $P^*$  and  $M^*$ . (Note that this is a different ‘tension’ from that which Kim believes occurs between mental and physical causes of the same physical event. For it is not assumed that  $P^*$  causes  $M^*$ . We shall return to the supposed tension between mental and physical causes of the same physical event.) I regard this sense of tension between causation and supervenience as bogus. I would like to scrutinize how Kim explains it.

Kim asks, ‘Where does the instance of  $M^*$  come from? How does  $M^*$  get instantiated on this occasion?’ He regards the two answers

(a) because by hypothesis  $M$  caused  $M^*$  to be instantiated,

and

(b) because, by supervenience,  $P^*$ , the physical supervenience base of  $M^*$ , is instantiated on this occasion,

as in ‘real tension’.

(a) and (b) are answers to different questions. One is a causal answer. The other simply cites  $M^*$ ’s supervenience base. Although Kim is aware of the difference between causation and supervenient determination, his formulation of the two answers runs together the different notions, misleadingly expressed by ‘because’ in the remainder of his discussion. I believe that if one keeps the difference steadily in view, the supposed tension itself comes to seem illusory.

Let us look at his attempts to bring out the tension:

I hope that you are like me in seeing a real tension between these two answers: Under the assumption of mind–body supervenience,  $M^*$  occurs because its supervenience base  $P^*$  occurs, and as long as  $P^*$  occurs,  $M^*$  must occur no matter what other events preceded this instance of  $M^*$ —in particular, regardless of whether or not an instance of  $M$  preceded it. This puts the claim of  $M$  to be a cause of  $M^*$  in jeopardy:  $P^*$  alone seems fully responsible for, and capable of accounting for, the occurrence of  $M^*$ . As long as  $P^*$ , or another base property of  $M^*$ , is present, that absolutely guarantees the presence of  $M^*$ , and unless such a base is there on this occasion,  $M^*$  can’t be there either.<sup>14</sup>

Assuming supervenience, if  $P^*$  occurs,  $M^*$  must occur, no matter whether the physical cause ( $P$ ) of  $P^*$  occurs or whether the mental cause ( $M$ ) of  $M^*$

<sup>14</sup> Kim, *Mind in a Physical World*, 42.

occurs. Unless the causal relation between an effect and its cause is metaphysically necessary, it is metaphysically possible that the effect, whether mental or physical, of a cause, whether mental or physical, could (metaphysically) have been instantiated, even though the cause were not. Most philosophers who think about these things believe that a given effect metaphysically could in many instances have had another cause. This point hardly puts ‘in jeopardy’ the claim that the given effect, whether mental or physical, has the cause, or causes, that it has. If the cause of a given effect *is* metaphysically necessary, then the point that the effect occurs regardless of whether the cause occurs is idle.<sup>15</sup>

To put the point another way: Kim writes that  $P^*$  alone ‘seems fully responsible for, and capable of accounting for, the occurrence of  $M^*$ ’. To be sure, on assumption of supervenience, and given that  $P^*$  occurs,  $M^*$  must occur. But  $P^*$  is not the cause of  $M^*$ .<sup>16</sup> Causes are not in general metaphysically sufficient for their effects. Moreover, unlike the cause of  $M^*$ , at least a portion of its supervenience base is simultaneous with  $M^*$ . Supervenience is a matter of how things hang together. Causation is a matter of how the things that hang together come about. Since  $P^*$  is not causally responsible for  $M^*$ , its role as supervenience base could hardly put the claim that  $M^*$  has a mental cause in jeopardy.

Kim’s formulation may encourage another mistake. Not only does his use of his notion of ‘responsibility’ encourage the idea that causation and the base–supervenient relation are in some way in competition for ‘responsibility’ for  $M^*$ . His notion of *accounting* encourages conflating these two relations with explanation.  $P^*$  surely does not explain  $M^*$  causally. Citing a physical supervenience base of  $M^*$  hardly gives a satisfying ‘account’ of its occurrence. It certainly does not obviate the need for a psychological explanatory account of  $M^*$ . Kim’s initial questions, which elicit answers (a) and (b), are unspecific in just the ways that invite the answers which tend to run together importantly different issues—causation, supervenience, explanation.

In a later book, Kim tries to bolster his claim that there is a ‘tension’ between base–supervenient determination and mental causation.<sup>17</sup> He counts this claim the ‘fundamental idea’ of his argument. Kim repeats his earlier point that as long as the supervenient base is in place, the mental occurrence must occur, no matter what happened before that occurrence. I think this point is ineffectual, for the reasons given two and three paragraphs back. Base–supervenient determination is not causation and cannot do its job.

<sup>15</sup> Note again that we are not yet discussing a supposed competition between physical and mental causal competitors. Here the supposed competition is between a mental cause of  $M^*$  and a physical supervenience base of  $M^*$ . My point here is that the supervenience base no more puts the mental cause of  $M^*$  in jeopardy than it puts its own physical cause,  $P$ , or a physical cause of  $M^*$ , in jeopardy.

<sup>16</sup> Kim makes this assumption also. Cf. *ibid.*, 44.

<sup>17</sup> Kim, *Physicalism, Or Something Near Enough*, 36–38.

Kim appeals to the theologian Jonathan Edwards as the first philosopher who saw the alleged tension. Kim attributes to Edwards the view that there is a tension between what he calls vertical determination and horizontal causation. He attributes the further view that vertical determination excludes horizontal causation. He calls this pair of views ‘Edwards’s Dictum’. Kim gives no evidence that Edwards was committed to Edwards’s Dictum.

Edwards held that there are no temporally persisting objects (hence, also no horizontal object-to-object causation) because he believed that

God is the sustaining cause of the created world at every instant of time. There are no persisting things because at every moment God creates, or recreates, the entire world *ex nihilo*—that is what it means to say that God is the sustaining cause of the world.<sup>18</sup>

Kim explains that on Edwards’s view, God’s being a sustaining cause renders causation between temporally successive events nugatory. For Edwards thinks that if God’s sustaining causation were withdrawn, the whole world would vanish, regardless of what had happened before. So putative other, preceding causes do not causally influence anything that comes after them.

Kim holds that it is ‘simple’ to see how Edwards’s Dictum, which Kim takes to have been illustrated in the theological doctrine, applies to the mind–body case, ‘causing trouble for mental causation’:

Mind–body supervenience, or the idea that the mental is physically ‘realized’—in fact, any serious doctrine of mind–body dependence will do—plays the role of vertical determination, or dependence, and mental causation, or any ‘higher-level’ causation is the horizontal causation at issue. The tension between vertical determination and horizontal causation, or the former’s threat to preempt and void the latter, has been, at least for me, at the heart of the worries about mental causation.<sup>19</sup>

Let us, for the sake of argument, take Edwards’s view to have full intuitive force. The view is that God’s causation preempts other putative causation. Edwards argues: if God’s sustaining causation were withdrawn, the preceding horizontal events would have occurred and the successive events would not have occurred. This argument tends to support the view that the preceding horizontal events are not really causes. Everything hinges on God’s sustaining causation. Nothing hinges on any putative causal power of the preceding horizontal events.

Kim takes two significant missteps in his use of Edwards’s views to support his own view that base–supervenience determination is in tension with mental causation (the first half of what he calls, misleadingly I think, ‘Edwards’s Dictum’.)

One misstep lies in his gloss on Edwards. Kim runs Edwards’s divine ‘vertical’ causation together with ‘vertical’ supervenience determination. Some of the intuitive force in Edwards’s position derives from God’s sustaining action’s

<sup>18</sup> Kim, *Physicalism, Or Something Near Enough*, 37. The words are Kim’s.

<sup>19</sup> *Ibid.* 38.

being *causal*. God's causation can seem to be in competition with ordinary causation. But base–supervenience determination is not causation. As I argued above, it cannot, even *prima facie*, take the place of causation in accounting for how things occur. In comparing Edwards's view about God's causation with his own view about supervenience determination, Kim has simply mixed apples with oranges—yet again.

Kim's second misstep is a failure to note the main source of Edwards's view that 'horizontal' causation between temporally successive events is illusory. Edwards's argument that God's causation preempts ordinary putative causation is based on a counterfactual claim that Kim nowhere replicates in his own account. If God's causation of the later event had been withdrawn, on Edwards's view, the antecedent event that is putatively the 'horizontal' cause would have existed, but would not have brought about any later event. No later event would have existed.

The supervenience case is disanalogous. Suppose that the physical supervenience base  $P^*$  of the mental effect  $M^*$  had not occurred. Then either  $M^*$  would have occurred with another supervenience base or it would not have occurred. If it did occur, there is no evident reason why  $M$  could not been its psychological cause. Suppose  $M^*$  did not occur. Then either  $M$  (with *its* supervenience base  $P$ ) did or did not occur. If  $M$  did not occur, it cannot be charged with being causally ineffective. If  $M$  and  $P$  did occur, there is no evident reason why they would not have been causally effective. In any normal counterfactual world, they would remain causally effective. Since they did not cause  $M^*$  and  $P^*$  respectively, there would be different causal conditions or laws, and they would have had other effects. None of these cases parallels Edwards's.

Edwards's argument depends on assuming that God can sustain the world to any given point and then withdraw support, depriving earlier events of any successors (hence of any effects). These assumptions about God's power have no analogs in ordinary counterfactual reasoning about physical and mental conditions in a supervenience relation. Edwards's view does not parallel Kim's. So Kim's attempt to use Edwards's view to bolster (or even illustrate) his claim that there is a tension between base–supervenient determination and mental causation is ineffective. It rests on compound conflation.

The deeper source of Kim's claiming tension where there is none is his arguing from metaphysical intuitions that abstract from what we know from causal explanation in science and common sense. Here we come again to the main the point of my original article: Our confidence in citing mental causes resides, or should reside, in psychological causal explanations. Nothing that Kim cites in his metaphysical discussion bears on psychological explanation. So nothing that he cites puts the hypothesis that  $M^*$  has a mental cause  $M$  'in jeopardy'.

Kim proposes to resolve the tension that he has purportedly identified by appealing to a new principle:

M caused  $M^*$  by causing  $P^*$ .

Kim thinks that it is justification enough of this claim that it helps resolve the alleged tension. He adds that ‘there may be’ a plausible principle that justifies this principle by entailing it:

To cause a supervenient property to be instantiated, you must cause its base property (or one of its [possible] base properties) to be instantiated.<sup>20</sup>

I see no reason to believe these principles. Causal explanations in psychology certainly attribute mind–body causation. They do so primarily where mental states or events cause bodily action-movement. We have no independent reason to think that all mental causes of mental effects cause physical effects that are supervenience bases of their mental effects. It is not independently evident—much less scientifically supported—that every mental event or property that causes a mental event, in every inference for example, causes some physical event or condition that is the supervenience base of its mental effect. Although one can construct a metaphysics that entails this claim, the claim is not supported by scientific explanation or intuitive reflection. I see no reason given so far to be any more favorable than agnostic about these principles.

If one reflects on what the supervenience bases of certain particular thoughts would be, supposing (as we are) that they have such bases, the principles seem wildly *implausible*. Take an occurrent thought that mercury occurs in Lake Baikal. Suppose that this thought is caused by prior thoughts and inferential steps. What is the supervenience base of this mental event?

I believe that I have given strong reasons to believe anti-individualism. I think that anti-individualism is in various ways supported or presupposed by relevant science. Some of the arguments for anti-individualism indicate that the supervenience bases of empirical thoughts are not local to the body of the individual. The supervenience base of a thought about mercury and Lake Baikal would involve a complex pattern of individual–environment relations, including causal-perceptual relations to objects in the environment. The supervenience base of such thoughts is a massively complex pattern that includes states in the thinker’s body at the time of the thought, but extends over large stretches of space and time. Such a pattern is not local to the individual’s body. The supervenience base of any belief or thought that might be a cause of the thought will also be complex, trans-temporal, and spread out in space.<sup>21</sup>

It would be to stretch matters to count such patterns as properties at all. They are surely not instances of any natural kinds. They are trans-temporal physical conditions, radically spread out in space as well as time. It is not plausible to

<sup>20</sup> Kim, *Physicalism, Or Something Near Enough*, 42, my insertion of the bracketed word. I do not understand how Kim understands the first principle if it is not already equivalent to the second. The first principle is evidently a generalization, not a claim about any particular mental or physical occurrences. And it is hard to see why Kim would think that it is merely contingent.

<sup>21</sup> Phenomenal properties may supervene on neural states. So this argument applies only to representational psychological states, but these are the states that are most prominent in causal psychological explanations.

think of them as causes of anything. The idea that the physical supervenience base of the thought is caused by the supervenience base of another mental state or event—or that a mental state or event causes such a supervenience base—has no intuitive or scientific plausibility.

I do not claim that it is incoherent to construct a metaphysics according to which there are physical–physical or mental–physical causal relations that take these sorts of supervenience bases as causal 'agents'. But I think that such a metaphysics has no claim on our belief. I see no explanatory potential in such a view. Both commonsense and scientific causal explanations take mental and physical causation to be spatially and temporally more local than any such metaphysics could allow. I believe that we have reason to dismiss such metaphysics. The relevant physical causation is to be understood in terms of causation among neural states. Neural states are not a supervenience base for at least many mental states.

Where does the argument go from this stage? Kim holds that any mental cause  $M$  of any mental effect  $M^*$  causes  $M^*$  by causing the physical supervenience base  $P^*$  of  $M^*$ . He also holds that  $P^*$  will have its own physical cause  $P$ . Then it seems that  $P^*$  has two causes— $M$  and  $P$ . He then holds that we can 'see reasons for taking  $P$  as preempting the claim of  $M$  as a cause of  $P^*$ .'<sup>22</sup>

The reasons that Kim cites in this particular argument are stated rather cursorily. He rejects the idea that the mental and physical causes of a given physical effect are to be taken as jointly 'sufficient' but individually 'insufficient' for their effects. He regards this view as incompatible with our understanding of both mind–body and physical causation. Kim further rejects the idea that the mental cause is sufficient for the physical effect and the physical cause is sufficient for the same physical effect. He rejects this idea by appealing to the physical causal closure principle.<sup>23</sup>

From these rejections he moves quickly to the view that mental causation is epiphenomenal, and the physical cause is doing all the causal 'work'. Since he believes that this is an undesirable conclusion, he does not accept it. His favored solution is to reduce all mental causal factors to physical factors, thereby purportedly restoring causal efficacy to the mental causal factors. I will return to the notion of 'work'.

I believe that the argument that Kim offers has too many difficulties to support his favored solution. I have not rejected supervenience, but I believe that its epistemic credentials are not strong. In what follows, I will ignore the 'broadly physicalist outlook' which drives so much of what is supposed to seem plausible. I will ignore the under-explained but apparently question-begging and overly strong principle of physical causal closure. I will ignore the implausible idea that

<sup>22</sup> Ibid. 43.

<sup>23</sup> Ibid. 44–45. There are elements in Kim's argumentation in these pages, in particular his appeal to a closest possible world argument to criticize overdetermination, that I find obscure and unsound. I will not discuss these elements in his argument.



mental causes cause mental effects always (or necessarily) by causing physical supervenience bases of those effects. I will ignore the idea that mental causes always cause physical effects.

I believe that the basic philosophical issues regarding mental causation are independent of assumptions about supervenience. What is at issue is, on the surface, relatively simple.

Psychological causal explanation indicates that mental events and mental properties are causally relevant. Sometimes mental–physical causation occurs. In such causation some mental events or states cause some physical event or state, where the mental aspects of the causes are causally efficacious.<sup>24</sup> Any physical event has a physical cause, with physical properties that are causally efficacious. So any physical effect of a mental cause will have a physical cause. Both mental and physical causes have properties that are causally relevant or efficacious. What is the relationship between the mental and physical causes (and their properties) in such cases?

There is a natural impulse to deal with the question by holding that the causes and properties are ‘the same’. This answer sounds good until one gets into the details of ‘sameness’. Token-identity claims, type-identity claims, claims that the mental is ‘made out of’ the physical, claims that mental explanation or properties are reducible to physical explanation or properties, claims that mental properties are second-order functional properties of physical properties—all have specific difficulties. I argued against the first in my paper. I think that the others bear no close relation to what we know from the sciences. Some of them fall into epiphenomenalism.<sup>25</sup>

<sup>24</sup> I need not take a position on how often mind–body causation occurs. It is enough that it is adverted to in a good bit of psychological explanation, and that the idea of mental causation being entirely confined to causing mental effects is almost as unacceptable as epiphenomenalism.

<sup>25</sup> I have not discussed Kim’s specific solution. This solution appeals to a reduction of psychological properties (at least those with intentionality) to functional properties, and then what he calls a further reduction of the latter properties, through reducing their realizations, to physical properties. I find his *notion* of reduction in this second stage very questionable. I also find the reduction of intentional psychological properties to functional properties a paradigmatically ungrounded philosophical claim. Analytic functionalism purports to give a conceptual analysis of psychological explanation into purely functional (causal role) terms. No plausible, or even specific, presentation of such an analysis has been given in a single case, for a specification of a single mental state-type, much less for all psychological discourse. The causal roles associated with specific psychological states like specific beliefs (beliefs like DNA contains a phosphate group or there is a red object on the distant hill) are too various to seem to admit such an analysis. So a specific account is needed. One cannot simply appeal to the general program. Moreover, the emptiness of specifications of causal roles makes the claim of conceptual equivalence incredible to disinterested reflection. To see this, reflect on even a schematic specification that uses only vocabulary that includes ‘causes’ and non-representational terms for stimuli and for behavioral response. No disinterested reflection will enable one to take seriously the idea that the specification has the same meaning or conceptual content as a mentalistic specification of a state.

Empirical functionalism takes the functional specification to yield not conceptual equivalence, but an account of the nature or constitution of representational psychological states. Empirical functionalism is, I think, no better off than analytic functionalism. Such a specification is far removed from any explanation that goes on in science. No science employs the functionalist theory envisioned by

I think a more exploratory, less committal metaphysical approach is more rational, and accords better with what we know. I think that our metaphysics is not yet very strong, epistemically, on these matters. What are strong are the claims that mind–body causation occurs, and that there are no gaps in the chains of body–body causation. Metaphysics should be pursued with a strong sense of its poor track record, and without writing as if metaphysical intuitions or principles are in a position to threaten mental causation, or put it in jeopardy.

I continue to advocate giving more attention to one largely unexplored line of inquiry. I doubt that there is any rational ground to think that a belief that a physical effect has both a mental cause and a physical cause forces a choice between maintaining that the causes are 'the same' and maintaining that they are 'in competition' or 'in tension'. I think that the credentials of any claim that such a choice is forced on us should be viewed much more critically. Alternative ways of understanding joint mental and physical causation invite more exploration.

In any case, I have so far found no ground to support the view that there is competition or tension between mental and physical causes. The explanation and motivation of the 'exclusion problem' from supervenience considerations are not persuasive. What other motivation might be marshaled?

Kim claims:

As long as each [the mental cause and the physical cause] claims to be a full cause of the event to be explained, a tension is created and we are entitled to ask, indeed compelled to ask, how the two purported causes are related to each other.<sup>26</sup>

We are certainly entitled to ask how the causes are related to one another. Let us also ask whether there is any antecedent reason to think that there is a tension, exclusion, or competition among physical and mental causes of the same physical event.<sup>27</sup>

functionalism, for example, a Ramsified functionalist theory. Any such theory would be devoid of any explanatory power. One cannot remove the theoretical terms (in this case, the mentalistic terms) from a scientific explanation and expect to have a comparable theoretical explanation. One cannot obtain equal, much less superior, power in giving causal psychological explanations by omitting the mentalist terms of the explanation in terms of blank descriptions of causal roles. Accounts of the nature of psychological states should illuminate the explanatory power of psychological explanation. (In fact, analytic functionalism should also illuminate the explanatory power of psychological explanation that it purports to analyze, but it fails to.) Again, not a single specific identification between mental properties and such functionalist properties has ever been carried through. The very idea that all of cognitive psychology can be reduced to some other theory that does not make any use of representational notions has no support in the way the relevant sciences are developing. Empirical functionalism is a tribute to the isolation of philosophy from scientific explanation. Both forms of functionalism seem to me to be waves of hands, without cognitive substance. As I indicated earlier, Kim's eventual position really leaves mental causation without any genuine causal role. For on his view, functional properties *per se* lack any genuine causal power. They free-ride on the underlying first-order physical causal powers. So Kim's eventual position falls into epiphenomenalism. See *Mind in a Physical World*, ch. 4.

<sup>26</sup> Ibid. 66.

<sup>27</sup> In this part of his discussion, Kim again misrepresents my views. In the first place, he argues again as if the issue were whether there is a place for metaphysical inquiry at all. As I have noted,

The quoted passage suggests that mental and physical causes ‘claim’ to be ‘full’ causes of the event to be explained. Causes do not make claims. Who makes this claim? Neither psychological nor physical explanation makes any claims about the other. The only sense in which we can reasonably say that the physical cause has a ‘claim’ to be the full cause is that it is causally sufficient. It is sufficient to make the physical effect occur. Citing it within a physical explanation is also sufficient to explain the physical effect in the terms of the physical explanation. Similarly, for the mental cause and the psychological explanation. The mental cause is causally sufficient to make the physical effect occur. Citing the mental cause within a psychological explanation is sufficient to explain the physical effect in the terms of the psychological explanation.

This situation leaves us with a question about how the causes are related, but it does not leave us with competition or tension. We can assume some sort of coordination or connection between the causes (insofar as they are distinct), since their causing the same effect is certainly not coincidental. There is no abstract compelling reason to think that this coordination or connection is just the causes being ‘the same’.

We assume, and the psychology tends to assume, that there is a physical cause of the physical effect. This is to say that the claim of ‘fullness’ by psychology does not exclude the physical explanation. Similarly, the physical explanation is certainly open to there being a mental cause, as long as it does not interfere with the physical explanation. The only reasonably grounded notions of ‘full cause’ are compatible with there being, in the relevant cases, coordinated, non-competing sufficient causes, mental and physical.

Kim also claims that the ‘exclusion problem’ arises ‘from the very notion of causal explanation and what strikes me as a perfectly intuitive and ordinary understanding of the causal relation’.<sup>28</sup> As I have just indicated, the different causal *explanations* do *not* seem to be mutually exclusive or in competition. No argument that Kim gives provides any non-question-begging ground to accept his view that there is competition among mental and physical *causes*.

There is, perhaps, an understanding of the causal relation from which intuitions like Kim’s about exclusion naturally arise. This is the understanding of

this was never the issue for me. In the second place, he quotes the paragraph in Section 3 of my paper, in which I claimed that it is ‘perverse’ to think that mentalistic explanation excludes or interferes with non-intentional (non-mentalist) explanation of physical movement. However, he uses ellipses to omit a key sentence in the paragraph. This omission makes it appear that I give a direct argument for my view that it is perverse to believe in exclusion or interference between mental and physical causes. The argument is made to appear to proceed directly from the premise that neither type of explanation (mental or physical) makes essential, specific assumptions about the other. Kim criticizes this argument. I did not give quite this argument. In fact, I use this premise only to support the view that ‘the relation between the entities appealed to in the different explanations cannot be read off the causal implications of either or both types of explanation’. This is just to say that sciences do not give us an account of a relation between mental and physical causes—hardly a controversial point. The perversity, in my view, derives from reflecting on the *particular* nature of the ‘very different types of inquiry’ embodied in psychological and physical explanation.

<sup>28</sup> *Ibid.* 66–67.

causation as occurring among physical events. In the physical domain, cases of different causes of the same outcome commonly fall into one of three categories. Either they are unusual cases of coincidental overdetermination. Or they are cases in which one causal factor is constituted of or otherwise 'resolvable' into the other.<sup>29</sup> Or they are cases in which the different causes are partial and in need of each other's physical contributions for their effect. None of these cases seems to fit mental causation.

In later work, Kim distills these claims about the causal relation into what he calls the Principle of Causal Exclusion:

If an event *e* has a sufficient cause *c* at *t*, no event at *t* distinct from *c* can be a cause of *e* (unless this is a genuine case of causal overdetermination).

Kim understands overdetermination to involve 'two or more separate and independent causal chains intersecting at a common effect'. The separate and independent clause suggests that causal overdetermination is a matter of a certain coincidence or accident. Kim and I would agree that overdetermination in this sense is unusual and that mental and physical causes of a physical effect are not cases of overdetermination in this sense.<sup>30</sup>

Kim takes the Causal Exclusion principle to be 'virtually an analytic truth with not much content'.<sup>31</sup> This strikes me as an amazing claim. It is surely a consequence of intuitions about causation guided by his metaphysical view. I think any notion of causation according to which the Principle of Causal Exclusion is virtually 'analytic' has lost touch with any notion of causation that is used in scientific or commonsense explanation.

Science certainly does not take mental and physical causation of a physical effect to exclude one another. Any such mental causation and physical causation are non-accidentally related. They are not independent. And hence they do not constitute a case of overdetermination, as Kim uses the term. There is certainly either a *prima facie* assumption that the mental and physical causes are distinct or at least an open question whether they are distinct. Nothing in commonsense or scientific usage supports the 'virtual analyticity' of the Principle of Causal Exclusion.

Kim holds that each science claims its cause to be 'sufficient' and not 'partial'.<sup>32</sup> He takes this claim to be evidence of competition. Our best understanding of sufficiency derives from the sufficiency of causal explanations within each

<sup>29</sup> Actually, the relations among different causes postulated in different physical sciences (say biology and physics, or geology and physics) are not understood in real depth. So this talk of constituency and resolvability is a wave of the hand. It is potentially misleading. I believe that the mind–body problem is much more difficult than the macro–micro physical problem. Even so, it would be a mistake to think that handling the latter problem is simple or straightforward.

<sup>30</sup> Kim, *Physicalism, Or Something Near Enough*, The principle is stated on p. 17. The construal of overdetermination is stated on p. 48. I think that the principle has serious difficulties in understanding even relations between instances of causation at different physical levels of explanation. Kim discusses this issue at some length on pp. 52–69. I find various aspects of this discussion unpersuasive, but I will not discuss these aspects here.

<sup>31</sup> *Ibid.* 51.

<sup>32</sup> *Ibid.* 53, 37–38.

science's domain. Neither physics nor psychology makes claims about the causes asserted by the other science. Insofar as a physical or mental event causes a physical event, it brings it about. In that sense, each cause is sufficient. I see no ground for either cause to be taken to exclude the other. We need to understand the relation while assuming that each cause has whatever sufficiency its causal power requires.

Kim asks, given that the physical cause is 'sufficient', what 'work' remains for the mental cause. It is not clear to me what makes this question seem forceful to him. I am inclined to think that the question trades on unclarified notions of sufficiency and work. The notion of work has homes both in physics and in talk of physical labor. The notion may elicit thinking of the psychological cause as like a further physical cause, offering an extra infusion of energy that is in fact not needed to supplement an already sufficient physical cause. Here mental causation would be implicitly regarded as a form of physical causation. In such a role it can easily seem to be an intrusive, competing, physical-like cause.

Alternatively, it may be that one can sustain a notion of work that is neutral as between mental and physical work. The idea of physical cause using up all the work so that there is no work left for the mental cause still seems to trade on a kind of hydraulic model. According to this model, so much energy is needed to get the job done. Given that enough energy is expended to get the job done by the physical cause, the mental cause is left without any need to expend its energy. This idea too seems to import a conception of the relation between physical and mental causation that is not sanctioned by ordinary explanations in the physical and human sciences. So it is reasonable to distrust the metaphor underlying the 'work' questions.

I believe that there is no strong, independent source of the idea that causation in psychology is in tension with causation in the natural sciences. The sense of tension is a product of assumptions of ungrounded metaphysics, or of metaphors which when held up against what we know, couched in literal terms, do not carry rational weight. The sense of tension is not an independently supported claim that a metaphysics is needed to explain.

If one asks, more neutrally, how the mental cause's causation is to be understood in relation to the physical cause's causation, we have, I think, a legitimate and unbiased question.

When we know from psychological explanation that mental events cause a physical event, where the mental properties are relevant to the causation, we also know from physiological explanation that there are physical events that cause the same physical event, where their physical properties are relevant to the causation. Each type of causal explanation is 'complete' (sufficient) on its own terms. But how are we to view the matter from a perspective that includes all the causal explanations and all the posited causal relations?

We assume that the mental cause could be effective only if there is a physical cause. Many are inclined to think that the physical cause would be effective regardless of whether there were a mental cause—sufficient to itself. This

'regardless' is problematic. It is true that the physical explanation does not need the psychological explanation to explain the caused physical event. However, the same is true of the psychological explanation. Psychological explanation need not appeal to an underlying physical story (which is largely unknown, in any detail, in actual cases) in order to give its causal explanation. We believe that the psychological causation could not occur without there being physical causation. It is hard to see how the particular physical causation could occur and there be no mental causation. There certainly remains an instinctive sense that the mental depends on the physical in a way that the physical does not depend on the mental. This sense is, I think, poorly understood.

The most common recent way to answer our question, 'How is the mental cause's causation is to be understood in relation to the physical cause's causation?', is to try to spell out in a satisfying way the idea that the causes are 'basically the same'. I think that this strategy has so far not yielded an account of 'basically the same' that provides a satisfying answer to the mind-body problem. And it is striking that among materialists there is no agreed-upon specific account of the 'sameness'. Another way to answer the question is to try to clarify the notion of mental causation in a way that accords with the idea that the mental and physical are not in competition and operate non-coincidentally and in concert. The two types of causation clearly operate together in some systematic way. What that way is remains to be understood. I believe that this second strategy has been under-explored.

The mind-body problem is the problem of understanding the relation between mental events and properties and physical events and properties. I think that there is no future in attempts to argue from an assumption about tension or exclusion to a conclusion about the mind-body relation. Our understanding of mental causation is no better than our understanding of the mind-body relation. The relation between mental and physical causation is not well understood, both because we have not solved the mind-body problem and because we do not have a satisfying understanding of mental causation, or indeed any causation. I include myself in this 'we'.

I think that an open attitude to exploring these matters is a better cognitive position than a metaphysics that assumes a vague generalized physicalism, and leans on visions of mind-body competition that are not grounded in anything that we know.

There is certainly reason to expect illumination from the progress of the sciences. At relatively primitive levels of the psychological, close connections have been established between psychological and neural causation. In understanding low-level vision, for example, the neural pathways and the functions of neural structures in visual processing are better understood than they were two decades ago. This knowledge should give us better tools for understanding mind-body relations. Even in these sorts of cases, a precise and satisfying characterization of the relations seems to me still to lie well ahead of us. Whether systematic, scientifically tractable correlations occur between neural pathways

and higher levels of perception and cognition is more doubtful, and certainly an open question.

The mind–body problem is difficult partly because the notion of causation itself is not well understood in philosophy. It is difficult partly because of the variety of ways in which psychological events and properties relate to physical events and properties. The problem is certainly not confined to understanding consciousness. It remains puzzling even for the representational aspects of mind.