

# In Defense of Epiphenomenalism

Jack C. Lyons

*Recent worries about possible epiphenomenalist consequences of nonreductive physicalism are misplaced, not, as many have argued, because nonreductive physicalism does not imply epiphenomenalism but because the epiphenomenalist implication is actually a virtue of the theory, rather than a vice. It is only by showing how certain kinds of mental properties are causally impotent that cognitive scientific explanations of mentality as we know them are possible.*

*Keywords: Functionalism; Epiphenomenalism; Nonreductive Physicalism; Type Materialism; Mental Causation; Reduction*

For several years, Jaegwon Kim and others have argued that there is a serious problem for functionalism and other nonreductive theories of mind (Kim, 1989, 1993, 1996, 1998, 2002; also, Kazez, 1994; Ludwig, 1994). The problem, we are told, is that nonreductive physicalism (NRP), in conjunction with certain intuitively plausible causal exclusion principles, implies a kind of epiphenomenalism.

Kim's now familiar argument goes roughly as follows. If mental events are not identical with physical events but, e.g., merely supervenient on them, then any given mental event and its corresponding physical event will compete for causal responsibility, and in such a competition the mental event is sure to lose. Any causal consequence of the physical event is, by a causal exclusion principle, *ipso facto* not a causal consequence of the mental event. Similarly, even if mental event tokens are identified with physical event tokens, a nonreductive theory denies that mental properties or types are identical with physical properties or types, and now another exclusion principle comes into play. If the physical properties of the event are causally responsible for its effects, then the mental properties are thereby causally excluded.

---

Correspondence to: Jack C. Lyons, Department of Philosophy, University of Arkansas, Fayetteville, AR 72701, USA. Email: jclyons@uark.edu

Call the first kind of epiphenomenalism “event epiphenomenalism” and the second kind “property epiphenomenalism”; the first claims that mental *events* are inefficacious, the second that mental *properties* are. Both kinds of epiphenomenalism are allegedly avoided by a reductive physicalism, which identifies psychological with physical properties. If psychological properties are identical with physical properties, then there is no room for competition and thus no room for epiphenomenalism.

A sizeable literature has grown up around this problem, and most of the authors working on this problem share Kim’s belief that epiphenomenalism of either sort would pose a serious objection to any view that implied it. I want to take a very different approach. I think that epiphenomenalism is not only not a problem for NRP, but is actually a *virtue*. More precisely, I think that there is a particular version of NRP which does imply property epiphenomenalism, but I think that this is a large part of what has always been attractive about the view. Thus, rather than endorse type materialism or work out a theory of causation that avoids exclusion principles, I think the proper response on the part of the nonreductive physicalist is to embrace, indeed to insist on, epiphenomenalism.

I begin by sketching in § 1 the version of NRP I will be assuming in what follows. I will not be concerned to argue for the view, nor to address any of the well-known objections except for the one that concerns epiphenomenalism. Next, in § 2, I try to clarify the relevant notion of epiphenomenalism and motivate the ensuing discussion by pointing out that this sort of epiphenomenalism may well be harder to avoid than is generally suspected, either by the defenders or the detractors of NRP. In § 3 I examine and reject the standard arguments against epiphenomenalism. In § 4 I describe what I take the virtues of (property) epiphenomenalism to be, and in § 5 I explain how the possibility of genuinely causal psychological laws is compatible with property epiphenomenalism.

## 1. Computationalist Functionalism

A standard version of NRP—in cognitive scientific circles, *the* standard version—has it that psychological states are functional states, in particular, computational states.<sup>1</sup> Classic statements of this general mind-as-the-software-of-the-brain view can be found in Block (1990b), Cummins (1989), Fodor (1975), and Pylyshyn (1984); it is implicit in such empirical work as Marr (1982) and Pinker (1997); a more recent philosophical articulation can be found in Harnish (2002). Though not necessarily the version intended by all the aforementioned authors, the version I have in mind and for which I will reserve the label ‘CF’ starts with (a) a generic computationalist functionalism, and further asserts (b) a token identity of psychological and physical states and (c) a roughly Fodorian (Fodor, 1974) view about the relation of the special sciences to physics.

Any functionalist theory of psychological types claims that a particular token is of the psychological type it is in virtue of its having a particular causal role. Of course, in order to preserve the multiple realizability that is the hallmark of functionalism,

only a proper subset of the whole causal role of a given state can enter into determining this type. A given brain event causes not only certain other brain events and bodily movements, but certain EEG readings as well; it has highly specific chemical and electromagnetic consequences. If *all* the causal consequences of a given state or event were relevant to the individuation of psychological types, multiple realizability would surely fail, for a differently realized token is bound to involve *some* difference in causal role.

CF is a computationalist theory in that it limits the *relevant* causes and effects (i.e., those that determine psychological type) to those causes and effects that are properly interpreted as computational.<sup>2</sup> Computation is just the syntax-driven manipulation of physically implemented symbols, though I intend the terms ‘syntax’ and ‘symbol’ to be construed broadly enough to include connectionism as a species of computationalism (see Harnish, 2002). CF is thus neutral regarding the classical–connectionist debate.

CF insists on token identity, thus precluding event epiphenomenalism; mental events will have whatever causal consequences the realizing physical events have. Thus the only kind of epiphenomenalism that could pose a threat to CF is property epiphenomenalism.<sup>3</sup>

CF’s commitment to a roughly Fodorian view about the special sciences (see especially Fodor, 1974) is most easily explained by invoking a terminological distinction between properties and kinds, according to which properties are plentiful, but kinds are harder to come by. Any disjunction of properties constitutes a property, whereas kinds are just those properties that are projected by some set of (true) laws.<sup>4</sup> CF holds that the set of pain realizers will be heterogeneous and unprojectible from the standpoint of physics in the sense that none of the properties projected by any laws of physics is coextensive with this set. *Pain*, therefore, would be a physical property but it would not be a physical *kind*. It could nonetheless be a psychological kind so long as there are true psychological laws ranging over pain.<sup>5</sup> The relationship between the laws of physics and psychology will receive more attention in § 5.

It is important to specify exactly what CF is a theory *of*. CF is (a) a theory about the relation between two scientific disciplines and how their respective kinds match up, and (b) a theory about the nature of the (cognitive) mind.<sup>6</sup> It is not, as are some theories in the philosophy of mind, a theory about our concept of the mind, or what we pretheoretically believe about the mind; it is about the mind itself, just as biology is not concerned with our beliefs about organisms and their histories but with actual organisms. To that end, CF takes current psychological theorizing to be at least roughly indicative of the true nature of the mind and tries to specify the relation between (true) physics or neuroscience and true, finished psychology, a discipline we perhaps know little about but which we can only assume will bear some resemblance to contemporary psychology. Of course, all bets are off if true neuroscience looks nothing like contemporary neuroscience or true psychology nothing like contemporary psychology.

‘Reduction’ is a much-abused term and means many things to many people, but the theory just sketched is nonreductive in a weak though very important sense.

By affirming multiple realizability, CF denies that the kinds invoked by psychology are identical with the kinds invoked by neuroscience or physics. Even this fairly weak nonreductivism suffices to supply CF with the standard purported virtues and vices; the theory is susceptible to Kim's worries about NRP as well as at least some of the autonomy claims that have traditionally motivated NRP.<sup>7</sup> If psychological kinds are not merely neuroscientific kinds, then psychological laws are not merely notational variants of physical or neuroscientific laws; the physical or neuroscientific counterpart to a psychological law will typically not itself be lawlike. With psychological laws thus distinct from neuroscientific laws, what counts as bad methodology for discovering the latter need not count as bad methodology for discovering the former. This is significant, for while it is clearly bad methodology to try to discover neuroscientific laws without examining brains, working cognitive psychologists virtually never examine brains but go on positing laws anyhow on the basis of behavioral data. This would be problematic if the laws they were positing were just neuroscientific laws in disguise.<sup>8</sup>

## 2. Property Epiphenomenalism

Because CF is incompatible with event epiphenomenalism, the only kind of epiphenomenalism that could be an issue is property epiphenomenalism. It is important, however, to be clear about exactly what this is. Recall how the anti-NRP argument is supposed to go. If psychological properties are distinct from physical properties, then an exclusion principle implies that if the psychological property is causally efficacious in any given instance, the corresponding physical property is not, and vice versa. But what does this mean? If properties are Platonic universals or collections of possibilia then they are *never* causally efficacious. If, as many believe, the only causal relata are events, and properties are not events, then again all properties are epiphenomenal. So property epiphenomenalism cannot be simply the claim that the properties are causally inefficacious. Nor can it be the claim that the instances of the property are inefficacious; this would be event epiphenomenalism, since the instances of the property of being in pain are pains (which according to CF *are* causally efficacious).<sup>9</sup>

Still, there is something to the notion of property epiphenomenalism; it makes perfect sense to say that the force of the collision caused the window to break, but the sound of the collision did not, or that the force, but not the sound, was causally relevant. Presumably, what is meant here is that the collision caused the window to break *in virtue of* its force but not in virtue of its sound.<sup>10</sup> Similarly, in the context of CF, property epiphenomenalism is the claim that even though a given mental state has causal consequences, it does not have them in virtue of being the kind of mental state it is. To use Horgan's (1989) term, the issue is one of "quasation," of whether psychological states have their effects *qua* psychological states.

I argue in §§ 3–4 below that property epiphenomenalism is not nearly as bad as has been generally supposed. The importance of this claim has a lot to do with whether,

as I suspect, such an epiphenomenalism is something we are all going to have to live with anyway. So before turning to the question of what is wrong—or right—with epiphenomenalism, I want to rehearse a few arguments that show how hard it is to avoid epiphenomenalism.

### 2.1. *Epiphenomenalism and Type Materialism*

One strategy for avoiding (property) epiphenomenalism is to reject NRP altogether in favor of a type materialism (TM), which holds that psychological kinds are identical with physical or neuroscientific kinds (Kim, 1996, 1998). Given the central role exclusion arguments have played in recent years, this is a natural move to consider. Embracing TM, however, makes it possible to avoid only some of the epiphenomenalism, not all of it. If the exclusion arguments are sound, they cut against Kim's own reductionism as well as NRP.

Recall that the key to the exclusion arguments is multiple realizability's implication of nonidentity. The competition for causal responsibility between the psychological and the physical kinds arises so long as the two are not identical. But even on the empirically dubious assumption that many important psychological kinds can be identified with the corresponding physical kinds, there will undoubtedly be many more psychological kinds that clearly cannot be thus identified. TM tends to focus only on what we might call the "type-properties" of mental state tokens. These are the properties of belonging to certain types, properties like being a pain, or being an edge detection, or being a belief that it's raining. But these are not the only, or even most interesting, properties of mental states. Other properties of mental states, like their intensity, their coherence with other mental states, their being held with conviction, and the like, are just as intuitively efficacious, and these properties are extremely unlikely to be identifiable with physical or neuroscientific kinds. Even if type-properties do reduce, epiphenomenalism will continue to threaten unless *all* psychological properties reduce.

Suppose for example that being in pain is type identical with being in brain state number 7547 (BS-7547).<sup>11</sup> This would make the property of being a pain identical with the property of being an instance of BS-7547, and an epiphenomenalism about this type-property would be avoided. But what about the other properties this token has, like being phenomenologically intense? Suppose, as is plausible, that the phenomenological intensity of a pain is realized in the firing rate of the neurons that code for pain: the higher the firing rate of these neurons, the more intense the pain. Still the property of phenomenological intensity is surely not identical with firing intensity. Some retinal photoreceptors fire more in the dark than when they are lit, so their firing rate correlates inversely with perceived intensity. Furthermore, such neural states are not consciously accessible, so the high firing could not have been *phenomenological* intensity anyhow. Thus, having a high firing rate is not the same property as being phenomenologically intense; the relation between the two in the case of pain is one of mere realization. So even if TM bars epiphenomenalism

regarding type-properties like being a pain, epiphenomenalism regarding properties like phenomenological intensity remains.

What we have here is the converse of multiple realizability. The functionalist typically makes a lot out of the many-to-one mapping from physical states to mental states, but there is also a many-to-one mapping from mental properties of mental states to physical properties. Where the mental can be identified with the physical, epiphenomenalism cannot threaten, but any many-to-one mapping blocks identification; multiple realizability is only one way to achieve nonidentity. Even if the only possible way to realize phenomenological intensity is in high neural firing rates, the properties of having a high firing rate and being phenomenologically intense are nonidentical. Consequently, someone like Kim, who argues from nonidentity to epiphenomenality, will be stuck by parity of reason with a fair amount of property epiphenomenalism himself. Although Kim can say that its being a pain has an effect, he must claim that its being phenomenologically intense doesn't. The high firing rate of the relevant neurons does have an effect; it causes the person to scream a little louder, display a stronger avoidance response, etc., but the pain's phenomenological intensity is something different and hence competing and so, by the exclusion arguments, must be inefficacious.<sup>12</sup> Surely Kim would not welcome such a result.

There are, of course, certain narrower properties Kim could cite, which are not shown by this argument to be epiphenomenal. For example, it may be possible to identify some physical property with the property of being a phenomenologically intense pain, thus saving the causal efficacy of *this* property. But note that this is not the same property as the property of being phenomenologically intense; the latter is a property that some pleasures have; the former is not. I am not claiming that all properties of mental states will be epiphenomenal, only that some will, and more, presumably, than Kim had intended.<sup>13</sup>

Similar problems beset the local reduction gambit. The ploy here is to reinterpret apparent cases of multiple realizability as cases of relatively narrowly conceived mental properties being singly realized. So the discovery that pain in humans is realized by a different kind of neural state than pain in octopi, we conclude not that pain is multiply realizable, but that pain-in-a-human is identical with BS-7547, while pain-in-an-octopus is identical with something else (Churchland, 1988; Kim, 1996, 1998; Lewis, 1969; Polger, 2004). Such a response, however, completely sidesteps the question of the causal efficacy of *pain*. In essence, local reductionism concedes that pain is multiply realizable and hence epiphenomenal, though we are encouraged to shift our attention from pain to pain-in-a-human, which is singly realized and thus efficacious.

As such, however, local reduction does nothing to secure the causal efficacy of *psychological kinds*. Insofar as the kinds involved in the laws of psychology are multiply realizable, local reduction has nothing to say about them. And because psychology does not attend to implementational details, its laws are typically not restricted in the way the local reductionist envisions. Psychology's laws usually range over such kinds as pain, rather than pain-in-a-human; hunger,

rather than hunger-in-cats; where this is not the case, the restriction is motivated by behavioral rather than implementational differences. Psycholinguists posit laws concerning language comprehension, not language-comprehension-in-a-certain-class-of-right-handed-males; this is true despite the knowledge that sex- and handedness-differences frequently have effects on such implementational issues as hemispheric asymmetry. To overlook this is to ignore the differences, in both method and type of laws posited, between cognitive psychology and cognitive neuroscience. Local reductionism might be an ambitious theory about what psychologists should be doing (*viz.*, cognitive neuroscience), but it cannot be a very plausible theory about what they are doing (psychology).

Although local reduction is officially silent about the status of psychological kinds, the basic motivating worries about multiple realizability imply a kind of epiphenomenalism—and likely an eliminativism—about psychological kinds. There is, on this view, such a thing as language-comprehension-in-a-certain-class-of-right-handed-male, but presumably no such thing as language comprehension full stop.<sup>14</sup> Since the former are not *psychological* kinds, affirming their existence and causal efficacy is no comfort to the friend of psychology. Local reduction thus does nothing to show that psychological kinds are causally efficacious.

## 2.2. Computationalist Functionalism and Epiphenomenalism

The other main line of response to epiphenomenalist objections against token physicalism is to develop a kind of causal compatibilism, whereby nonidentical properties do not *ipso facto* compete for causal responsibility (e.g., Fodor, 1989; Horgan, 2001; LePore & Loewer, 1987; Loewer, 2002). As important as such work has been in advancing our understanding of causation, and as sympathetic as I am to the idea of causal compatibilism, this response does not actually secure mental efficacy, but only staves off one challenge to it, perhaps not even the most serious of such challenges. CF, however, implies property epiphenomenalism whether we accept any causal exclusion principles or not, and so will similar versions of NRP.

An initial argument linking CF to epiphenomenalism starts with the observation that psychological laws individuate states on the basis of their semantic properties. However, we seem to lack any naturalistic theory of content on which semantic properties can be anything but epiphenomenal. According to one prominent theory of content, conceptual role semantics, mental state tokens derive their meaning from the causal role they play in inference. If this is the case, then they have the meaning they do in virtue of having the causal role they do, not vice versa (Cummins, 1992); i.e., they do not have the causal role they do in virtue of having the content they do. Things are little different if we suppose an indicator semantics, which claims roughly that representations mean what they do in virtue of the external stimulus that reliably causes them. Since it is always possible to vary the external world without varying the internal construction of the agent, in good Twin Earth fashion, and since the internal construction of the agent that will determine at least some (aspects of) behavior,

content is epiphenomenal with respect to such behavior on this view as well (Ludwig, 1994). Even if we assume some kind of picture theory (Cummins, 1996), there is no reason to suppose that content *per se* is doing any work; instead it is the structural character of the representation that determines both its content and its causal role. Content and causal role are related as something like joint effects of a common cause, rather than the former being a cause of the latter (though of course, structure *produces* content and causal role without actually *causing* them). Causal role is not the result of content; both are the result of something else.

Additionally, CF, as a species of computationalism, explicitly views the mind as involving a “syntactic engine driving a semantic engine” (Block, 1995; Haugeland, 1981). This also brings with it a kind of property epiphenomenalism (Kim, 1996). The instantiation of psychological properties depends on the instantiation of syntactic properties, not the other way around. A token is of the psychological type it is in virtue of being of the syntactic type it is, and not vice versa.

Finally, the most general and, I think, compelling reason to think that CF is committed to property epiphenomenalism is simply the fact that CF is a version of functionalism and epiphenomenalism is already inherent in the basic statement of functionalism. The standard rallying slogan of functionalists is that “it is what it is because it does what it does.” The *because* relation, however, is asymmetric. If  $x$  is true because  $y$  is true, then it cannot be the case that  $y$  is true because  $x$  is true. Similarly, if  $x$  obtains in virtue of  $y$ 's obtaining, then it is false that  $y$  obtains in virtue of  $x$ 's obtaining.<sup>15</sup>

Strictly speaking, a theory might count as functionalist merely by claiming that a given token is of a particular psychological type *if and only if* it has a certain functional role, without actually claiming that the role *makes for* the psychological type. The kind of functionalism I have in mind here, however—and I think this is how the view has always been intended—purports to actually *explain* psychological type in terms of causal role. It is not merely that it is what it is if and only if it does what it does; it is what it is *because* it does what it does. Causal role is metaphysically and therefore explanatorily prior to psychological status, and so being of the psychological kind it is cannot be the reason why a given token has the causal consequences it does.<sup>16</sup> One could, perhaps, claim that this thing holds doors open because it is a doorstop, but this is not the route that functionalism takes; instead, functionalism claims that it is a doorstop because it holds doors open.<sup>17</sup> This is not just an oversight on the part of functionalism; as I argue in §4, it is an essential component of the best versions of the theory and a large part of what makes the theory so attractive.

This argument rests on an asymmetry principle (or family of principles): if  $x$  because  $y$ , then  $\sim(y$  because  $x)$ ; if  $x$  in virtue of  $y$ , then  $\sim(y$  in virtue of  $x)$ ; if  $x$  is prior to  $y$ , then  $y$  is not prior to  $x$ . These all seem clear enough and quite obviously true. The only version of the causal exclusion principle that is *obviously* true is the one that claims that in cases of non-overdetermination, there are not two or more *independent* causes of a given event. But of course, the allegedly competing mental and physical properties were never claimed to be independent. The higher-level property is

exemplified because the lower-level one is. The asymmetry principle is considerably weaker and more plausible than the relevant forms of the causal exclusion principle.

If having the causal role it does determines and is thus prior to a state's psychological type status, then the state cannot have its causal consequences in virtue of being of that psychological type. Instead it must be the other way around: it is of that type (partly) in virtue of having the causal consequences it does. Thus, no appeal to controversial theses about causal exclusion is necessary to pin property epiphenomenalism on the functionalist. This kind of epiphenomenalism was built into functionalism from the very beginning.

### *2.3. The Ubiquity of Epiphenomenalism*

Epiphenomenalism thus seems to be more prevalent than suspected by either the friends or foes of NRP. The two main options for a token physicalist who wants to avoid epiphenomenalism have been type materialism and causal compatibilism, but neither alleviates all the worries concerning epiphenomenalism. I do not claim that the above arguments are conclusive, and some are quite familiar, but they do serve to make a *prima facie* case for a link between property epiphenomenalism and physicalism. The standard attempts to accommodate the efficacy of mental properties do not succeed.

Despite the fact that these arguments are nondemonstrative, they do strongly suggest that we all have to live with property epiphenomenalism. If so, then perhaps the best theory is the one that can best make a virtue of necessity by getting the epiphenomenalism to do some theoretical work for it. This motivates an alternative approach, one that does not require working out a compatibilist theory of causation or a reductionist theory of psychological types.

## **3. Why Not Epiphenomenalism?**

Since the approach I will recommend is one that embraces property epiphenomenalism, it is helpful to remember why we were supposed to be opposed to epiphenomenalism in the first place. Why was epiphenomenalism such a bad thing again?

### *3.1. Intuitive Objections to Epiphenomenalism*

It is crucial to reiterate the distinction between event epiphenomenalism and property epiphenomenalism. Though CF implies that property epiphenomenalism is true, it implies that event epiphenomenalism is false. Unfortunately, the literature is full of objections to epiphenomenalism that focus entirely on event epiphenomenalism. The following is typical:

First, the possibility of human agency evidently requires that our mental states—our beliefs, desires, and intentions—have causal effects in the physical world . . . Second, the possibility of human knowledge presupposes . . . the causation of perceptual experiences and beliefs by physical objects and events around us.

Reasoning . . . involves the causation of a new belief by an old belief . . . (Kim, 1998, p. 31)

Nowhere in any of this is there any reason to deny *property* epiphenomenalism; the case here is at most one of guilt by association with event epiphenomenalism.

Perhaps it is simply supposed to be intuitively obvious that property epiphenomenalism is false. Horgan (1989), for example, claims that property epiphenomenalism is “hardly less offensive to common sense” than event epiphenomenalism (p. 47). Our intuitions, however, seem to be far more strongly opposed to event epiphenomenalism than to property epiphenomenalism. It is clear that we are pretheoretically convinced that mental states have consequences, but it is not clear that we were convinced that they have these consequences in virtue of being the kind of mental state they are.

More importantly, however, it is questionable just how much weight such pretheoretic intuitions should carry. CF is a posttheoretic hypothesis about the mind, not about our commonsense beliefs about the mind or our concept of mind. We would never dream of holding high-level theories in physics hostage to folk physics; why should psychology be any different? Despite a long tradition of *a priori* philosophy and introspectionist psychology, CF refuses to take any nonempirical view of the mind overly seriously. Prosopagnosia, movement blindness, physicalism itself: all violate our commonsense, folk psychological expectations. Much about the mind is counterintuitive. If the best empirical work in psychology all presupposes a theoretical foundation that implies property epiphenomenalism, then intuitions to the contrary should simply be retrained. If functionalism turns common sense on its head by making causal role prior to psychological type, we should simply conclude that common sense was wrong once more. I do not deny that *a priori* intuitions have a place in the philosophy of mind, or even in science, but anti-epiphenomenalist intuitions are intuitions about contingent, presumably empirical, matters of causation and thus among the least secure intuitions we have.

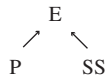
There is a different kind of intuitive case against property epiphenomenalism urged by, e.g., Jerry Fodor (1989). If psychological properties are epiphenomenal, then so are all special science properties, and it is these that are intuitively efficacious. Fodor’s primary examples are mountainhood and airfoilhood:

Untutored intuition suggests, it is because Mt. Everest is a mountain that Mt. Everest has glaciers on it top; and it is because Mt. Everest is a mountain that it casts such a long shadow; and it is because Mt. Everest is a mountain that so many people try to climb Mt. Everest. (p. 61).

Again, our intuitions are far from clear on this point; my own intuitions tell me that it is Mt. Everest’s *height* that is responsible for its having glaciers and casting long shadows, not its mountainhood *per se*. (And it is their belief that it’s a mountain, not its actual mountainhood, that causes people to try to climb it.)

But what about airfoilhood?

It would be quite mad to say that *being an airfoil* is causally inert. Airplanes fall down when you take their wings off; and sailboats come to a stop when you take



**Figure 1.** The Relations Among Physical Properties (P), Special Science Properties (SS), and an Event’s Complex of Effects (E), According to Causal Compatibilism.



**Figure 2.** The Correct Picture of These Relations.

down their sails. Everybody who isn’t a philosopher agrees that these and other such facts are explained by the story about lift being generated by causal interactions between the airfoil and the medium. If that *isn’t* the right explanation, what keeps the plane up? If that *is* the right explanation, how could it be that *being an airfoil* is causally inert? (Fodor, 1989, p. 62)

Again what we have is an argument against event epiphenomenalism, not an argument against property epiphenomenalism. What Fodor’s explanatory sketch here indicates is that airfoils are causally efficacious. His conclusion, that *airfoilhood* is efficacious, is unsupported.

Nor is there anything “quite mad” about claiming that airfoils have causal powers, though not in virtue of being airfoils. Following the Mt. Everest example, they may very well have their causal powers in virtue of their shape, texture, rigidity, and the like. Additionally, we can borrow an argument from §2 to show that they do not have their causal powers in virtue of being airfoils. Airfoilhood is a functional property; therefore, the wing of a plane is (or realizes) an airfoil in virtue of generating lift, not the other way around.

In diagram form, the causal compatibilists see the special scientific (SS) and the physical (P) properties of some event as standing in a causal relation ( $\rightarrow$ ) to that event’s complex of effects (E), as in Figure 1.

On this view, both P and SS are prior to E, and the kind of priority involved is a causal priority. The view I am urging has P being causally prior to E but has E standing in a constitutive relation ( $\Rightarrow$ ) to SS, as in Figure 2.

The priority relation between SS and E is reversed here, and it is a different kind of priority. Whereas causal compatibilism claims that special scientific properties are causally prior to the complex of effects, I hold that the complex of effects is metaphysically prior to the instantiation of the special science properties.

### 3.2. Explaining Away Anti-Epiphenomenalist Intuitions

Surely, however, our intuitions require causally efficacious properties. There is, after all, an obvious difference between such intuitively irrelevant properties as the color

of the ball that breaks the window, on the one hand, and the sail's being an airfoil, on the other. I admit that there is a difference here, though it is not one that needs to be cashed out in terms of the causal efficacy of the respective properties. Our intuitions about the causal relevance of airfoilhood likely result from our tacit recognition of a kind of counterfactual dependence of lift on airfoilhood. We can say that *A* counterfactually depends on *B* just in case if *B* had not obtained, *A* would not have. Lift thus counterfactually depends on airfoilhood in that if planes' wings did not realize airfoils, they would not generate lift. The case is different with the ball breaking the window, as the window breaking does not counterfactually depend on the color of the ball.

Though all this is true, it does not begin to imply that lift *causally* depends on airfoilhood, i.e., that the property of being an airfoil is causally efficacious, partly because none of this implicates *airfoilhood*, rather than airfoils, but also because counterfactual dependence is notoriously easier to come by than causal dependence. We might just as well claim that one's not marrying causally depends on his bachelorhood, on the grounds that if there were no bachelors, all men would have married. Such claims not only get the priority relation reversed but confuse causal priority with metaphysical priority. In fact, the so-called counterfactual dependence of lift on airfoilhood is not even happily viewed as genuine dependence, since (a) it is not asymmetric, and (b) the generation of lift is prior to the airfoilhood, rather than the other way around.

Nor will it matter much if we turn our attention to more sophisticated counterfactual accounts of causal relevance. Any adequate account must avoid making bachelorhood causally relevant to not marrying, but the very sophistications that render bachelorhood inert will have the same results for functional properties. Being unmarried is simply part of what it is to be a bachelor and so the latter cannot be causally relevant to the former. Similarly, for any effect which enters into the functional specification of a functional property, that effect's counterfactual dependence on the instantiation of the functional property is not a matter of causal relevance, for having such an effect is simply part of what it is to have that functional property.

To take a concrete example, consider the theory of causation Loewer (2002) endorses. According to Loewer, *M* causes *E* iff (i)  $\sim M \gg \sim E$  (i.e., if *M* were not the case, *E* would not be), and (ii) there is no *N* which preempts *M* with respect to *E* and is not preempted by *M*. *N* preempts *M* with respect to *E* iff (a)  $\sim N \gg \sim E$  and (b)  $(N \ \& \ \sim M) \gg E$ . My being unmarried counterfactually depends on my being a bachelor, though nothing could preempt that, since nothing would, in conjunction with my not being a bachelor, counterfactually imply my being unmarried. Hence, there is no *N* that would allow (b) to be satisfied; therefore, (ii) is satisfied, so according to Loewer's view, my being a bachelor causes me to be unmarried. The way to fix this, of course, is to require that *M* not entail *E*. (*N* must be similarly restricted, on pain of any candidate cause being preempted by some noncausal factor.) Earlier formulations with LePore (LePore & Loewer, 1987, 1989) do explicitly require that *M* and *E* be logically and metaphysically independent, and perhaps an additional

requirement Loewer (2002, p. 660) mentions in a footnote is intended to effect the same result. However, once such a clause is added, the view no longer rebuts epiphenomenalism, since the instantiation of a given mental property, according to functionalism, does entail at least some effects. In fact, the modified causal theory, with its additional requirements, actually implies epiphenomenalism (see also Ludwig, 1994). *M* cannot cause *E* unless *M* and *E* are metaphysically independent, which, according to functionalism, they are not.<sup>18</sup>

It is generally recognized that counterfactual dependence is weaker than causal dependence. What is not generally recognized is that this fact undermines intuitive arguments against epiphenomenalism, for the relevant intuitions may be about counterfactual, rather than causal, dependence. Those who have the most invested in the falsehood of property epiphenomenalism, the proponents of causal exclusion arguments, should be especially interested in this way of undercutting our anti-(property-)epiphenomenalist intuitions. Causal exclusion arguments fail to even get off the ground without presupposing some notion of causation more robust than mere counterfactual dependence, since counterfactual dependence is clearly not exclusive. If there is that much more to causation, however, there must be even that much less to any intuition of causal relevance that is based on an intuitive awareness of counterfactual dependence.

Thus, I think that the standard objections to property epiphenomenalism fail. Many simply evade the real target and attack event epiphenomenalism instead. The others are unconvincing in their own right. However, I want to claim not only that property epiphenomenalism is not a crippling vice for CF, but that it is actually a virtue. The fact that CF makes mental properties epiphenomenal is a large part of what is so attractive about CF.

#### **4. Why Epiphenomenalism?**

The case for a generic materialism has always been largely empirical: traditional problems for dualism involve consonance with physical conservation laws and difficulties in explaining the known brain-behavior dependencies. But extraempirical considerations, including methodological principles like Ockham's razor, have always been equally important. Dualism is sufficiently unparsimonious that if our only options were dualism and eliminativism, the latter might very well be the better choice. Better perhaps to do away with mentality altogether than to countenance it as something fundamental and novel. Noneliminative materialism tries to find a comfortable halfway position: mentality is real, but not quite as real as Descartes thought it was. We can avoid at least the grosser violations of Ockham's razor by showing how mental states and properties can be reduced to, or supervene on, or somehow be cashed out in terms of something more fundamental. As Fodor (1987) says of content elsewhere, so too of mentality more generally: if it is real, it must be really something else.

The central idea is that we want mentality to turn out to be real, but not *too* real. The virtue of property epiphenomenalism is that it ensures that mentality will not be too real. (The burden of § 5 will be to show how it can nonetheless be real enough.) But what does this mean, and why is it desirable?

#### 4.1. Parsimony

One need not be a neo-Platonist to find something coherent behind such vague talk about different degrees of reality. It is not, of course, that anything is literally more real than anything else but that some things are more basic, more fundamental than others. Psychological kinds are less fundamental than neurological kinds, which are in this metaphorical sense less real than physical kinds, etc. Physical kinds, or at least some of them, are truly fundamental. Mass, charge, and the like are basic features of the world and not features that something has in virtue of anything else. This underwrites the standard picture of nature as hierarchically organized. The very term ‘reduction’ suggests something like this; to reduce something is usually to reduce it to something more basic, more fundamental. Unfortunately, the theory of mind most strongly associated with ‘reduction’ is type materialism (TM), and TM does not quite offer a reduction in this sense. While CF claims that psychological kinds are realized by physical kinds and that the latter are more fundamental, such a move is blocked for TM, which identifies psychological kinds with physical kinds. There is a sense in which mental kinds cannot be *reduced* to physical kinds, because mental kinds *just are* physical kinds.

Proper use of the term ‘reduction’ aside, the point is that CF, but not TM, can claim that physical properties are more fundamental than psychological properties. If *F*-ness and *G*-ness are the same property, then clearly neither is more fundamental than the other. Such a view produces a number of difficulties that will soon command our attention, but for the moment, notice that TM thereby makes psychological properties as fundamental as the corresponding physical properties. Not only is this counterintuitive in its own right, but the type physicalist will have a hard time explaining how psychology manages to be less fundamental than the relevant parts of physics.<sup>19</sup> TM can still hold that physical *concepts* are more fundamental than psychological *concepts*, but even if this is true, it only gives us an epistemic hierarchy, not the metaphysical hierarchy we wanted.

Of course, by refusing to identify psychological kinds with the physical kinds that realize them, CF commits itself to the existence of more kinds than does TM. This may seem to put CF at a disadvantage vis-à-vis Ockham’s razor. Whether or not this is so, however, depends on how Ockham’s razor is formulated, for the epiphenomenalism inherent in CF indicates that it is not committed to any more *causally efficacious* kinds than is TM. Suppose CF and TM both admit the existence of physical properties  $P_1$  and  $P_2$ ; they also admit the existence of psychological property  $M$ , which TM identifies with  $P_1$  and which CF claims can be realized by  $P_2$  as well as  $P_1$ . TM recognizes three properties:  $P_1$  (a.k.a.  $M$ ),  $P_2$ , and the disjunction  $P_1 \vee P_2$ , only the first two of which are causally efficacious.

CF recognizes four properties:  $P_1$ ,  $P_2$ ,  $P_1 \vee P_2$ , and  $M$ , again only the first two of which are causally efficacious.<sup>20</sup> Ockham's prohibition against the positing of novel entities does not obviously apply to causally inert ones. There clearly are such "Cambridge properties" as the property of being prime or identical with Bertrand Russell; it is just that such properties do not make any difference in the world. In some sense that is difficult to articulate fully, commitment to such properties should be distinguished from commitment to phlogiston and the like; the former are mere constructs and are intended as such. But it is precisely this feature of them that makes their inclusion in an ontology harmless. Whether Ockham's razor favors TM over CF depends in part on whether the principle applies to all properties or only purportedly efficacious ones.

#### 4.2. Explanatory Power

It also depends on whether and to what extent CF is able to derive explanatory power from these additional posits. Ockham's razor is not merely the injunction to keep one's ontology small, but to refrain from enlarging it *without explanatory advantage*. The principle itself says nothing about how to weigh increased explanatory power against inflated ontology, and here different philosophers seem to have different opinions. Some think a large loss in ontological parsimony is a fair price for a slight increase in explanation, others that a great increase in explanation is needed to justify a slight inflation of ontology. I myself favor Ockham's razor because explanation is clearly an important goal, and the more inflated one's ontology is, the more the unnecessary entities tend to obscure rather than elucidate. The claim that angels push the planets around in their orbits raises far more questions than it answers. On this view, ontological parsimony is not necessarily a good in and of itself, but is only good because it is conducive to the development of the most comprehensive and satisfactory explanations.

In any case, explanatory power is *an* important virtue, and it is here that the chief merit of epiphenomenalism lies. CF is explanatorily superior to any other theory of mind, and this is true largely because of the epiphenomenalism implicit in it.

If mental states are computational states, then we are assured that an entirely physical thing can have them. Even more importantly, however, we can understand *how* an entirely physical thing can have them. Computers and brains are, at bottom, completely unintelligent; they are merely physically complex enough to engage in complicated internal interactions. Certain of these physical interactions can, at a suitable level of abstraction, be viewed as the formal manipulation of symbols, provided that the right sort of mapping exists from the causal interactions among physical states to the transformations among symbols. Since there is as yet no semantics in the picture, there is no intelligence in the picture, and there is nothing here more mysterious than the physics of the system: voltage-driven gate switching, wheel turning, propagation of neural signals, etc. To say that the symbol manipulation is formal is in part to say that it is entirely mechanical in the sense

that it requires no knowledge of the domain, or indeed any intelligence whatsoever. Now if the right sort of mapping (presumably an isomorphism; see Block, 1995; Cummins, 1989) exists between this symbol manipulation and some semantic domain, the symbol manipulation can be given a semantic interpretation. A few well-known complications aside, this is all there is to mentality; if something is semantically interpretable then it is a mental state.<sup>21</sup>

Systems whose state transitions are subject to consistent semantic interpretation are bound to be relatively rare, for very few things will ever have internal states that bear the right causal relations to each other unless specifically designed to do so, either by their programmers in the case of machines, by evolution in the case of brains, or by earlier stages of themselves in cases of learning. We explain the causal structures of the systems by appeal to just these factors of programming, evolution, and self-modification (the basis of which must itself result from either programming or evolution). But a different kind of explanation is on the table for explanations of intelligent behavior *qua* intelligent behavior. How does a dumb, physical device manage to do something as intelligent as playing chess? *By doing something else*, something that requires no intelligence. What the system is *really* doing is blindly though uniformly manipulating symbols that mean nothing to it; intelligent behavior, indeed intelligence itself, is an inert side effect of a far less mysterious phenomenon. Mentality, on this view, is a mere byproduct that has no independent contribution to make to the causal powers of the implementing mechanisms.

We explain mentality by showing how a system can be at once intelligent and unintelligent. What the system is doing is at one level blind and mechanical, at another level creative and rational. Such an explanation works, however, only if the levels (the physical or neural on the one hand, and the psychological on the other) and the behaviors exhibited at these levels, are genuinely distinct. For anything like this explanatory strategy to work, the physical properties must be not only distinct from the psychological properties, but actually contrary to them. The property of being a symbol writing must have the property of being unintelligent, while the property of being a chess move must have the property of being intelligent. To explain the “ghost in the machine” we need to maintain in some sense the ghost, the machine, and a distinction between the two (though we won’t view the ghost as immaterial, of course), but the ghost is, according to CF, not a mover of the machine but a causally inconsequential side effect of the machine’s operation.

It is this largely deflationary account of psychological phenomena that allows CF to explain mentality. Taking the functionalism and computationalism as read, the story just bruited takes the mystery away from the mental in great part by taking the efficacy away. The account is explanatory *because* it is deflationary, and its being deflationary consists of its making mentality epiphenomenal. The epiphenomenality of psychological properties ensures that there will be no further explanatory work to be done, once the less mysterious physical phenomena have been accounted for.

### 4.3. CF Explanation

There are three key features of the explanatory strategy pursued by CF that make it so attractive:

1. It offers what we might call “instantiation explanations”: explanations of why some higher-level property is instantiated, by appeal to the instantiation of a lower-level property.
2. It offers “realization explanations”: explanations of how the lower-level property realizes the higher-level property.
3. It accomplishes both of these by maintaining a distinction between higher-level and lower-level properties and insisting that the latter are doing all the work.

Endorsing TM, according to which the higher-level properties just *are* the lower-level properties, would render the higher-level properties efficacious at the cost of trivializing (1–3).

It is an essential part of the CF story that the chess machine manages to perform the intelligent task of putting an opponent in check *by means of* performing the unintelligent task of blindly flipping switches. This only makes sense if the intelligent task is not the same as the unintelligent task and if the latter is somehow prior to the former. The only way to have a single process token be both smart and stupid at the same time is by having it be an instance of two different types, e.g., the smart type *putting one's opponent into check* as well as of the dumb type *flipping such-and-such switches*. Thus, a single token instantiates two distinct properties, one intelligent but epiphenomenal and the other efficacious but unintelligent.

If putting an opponent into check were really identical with flipping such-and-such switches, then the system could not do the former *by means of* doing the latter. If *F-ness really is G-ness*, then something's having *F-ness* cannot be prior in any real sense to its having *G-ness*. Similarly, though TM can say that mental properties have no causal contribution to make over and above the causal powers of the physical properties, it cannot mean anything nontrivial by this claim. For it is just the claim that *F-ness* has no causal contribution to make over and above that of *F-ness*, which is the unilluminating claim that *F-ness* has no causal powers that it does not have.

Nor can an identity theory like TM offer full-blooded instantiation explanations. An instantiation explanation explains why there is something rather than nothing in the domain in question. One cannot explain the existence of water by pointing out that H<sub>2</sub>O exists, since the two are identical. There are historical, geological, and cosmological explanations for the existence of water, but no instantiation explanations. To treat the existence of H<sub>2</sub>O as explaining the existence of water would be treating the existence of water as *self-explanatory*, which it patently is not. It does seem, however, that instantiation explanations are appropriate for mental properties. Why are there pains, i.e., why do some things instantiate pain? In addition to historical and evolutionary explanations, instantiation explanations are also available: there are pains because there are instances of BS-7547.

TM's situation is even worse with regard to realization explanations. Necessary identities—and property identities are always necessary—are inexplicable. We can explain why Franklin is identical with the inventor of bifocals, but we cannot explain why Everest is Gaurisanker or why being a bachelor is being an unmarried man. Nor, if being in pain just is being in BS-7547, can we explain why or how this is true. Consequently, there would be no substantive explanation for why brains, but not, say, livers, realize mental properties. Where CF would point out the difference in causal structure between the two organs, TM can merely note that livers are never in *brain* state number 7547.

Christopher Hill (1992) recognizes that TM offers no realization explanations but claims that since the identities are necessary, there's nothing that needs to be explained. The problem with this, however, is that there clearly *is* something that needs to be explained; insisting that the question does not arise does not prevent the question from arising. Questions about how the brain realizes various kinds of mentality are among the most important and pressing questions we have asked, and the fact that we have the beginnings of informative answers to them testifies to their legitimacy. CF can explain, for example, why this neuron firing realizes an edge detection: because the firing of this neuron plays the same role at the neural level that the edge detection plays at the psychological level. Consequently, the fact that this neuron is firing explains why there is at least one mental state. So CF enables both realization and instantiation explanations. If our goal is to explain as much as possible, this ought to count heavily toward CF.

Everyone has to admit some brute facts; the difference is that for CF, the facts in question are conceptual/stipulative facts about the individuation conditions for mental states. For TM, on the other hand, the brute facts are empirical facts about the relationship between the mind and the body. The facts that turn out to be brute on this theory are just those that are most in need of explanation. Eliminativism would be preferable, for it would at least explain why these apparently legitimate questions are in fact not. We can reduce puzzlement either by explaining the phenomena or by showing that the putative phenomena are not in fact real, but not by insisting that the *puzzlement* isn't real.

#### 4.4. *Functional Reduction?*

Kim has done much to obscure all of this in some recent efforts to have his cake and eat it too (Kim, 1998, 2002). He wants to maintain TM to ensure the causal efficacy of psychological properties and a functionalist theory of reduction to retain realization explanations.

A functional reduction of a property works like this: (step i) the property must first be construed as a second-order functional property through a *functional definition*, a definition in terms of its "causal role"; (step ii) we must then find its "realizers," that is, first-order properties that fit the causal role specified by the functional definition; and (step iii) we must have an explanatory theory that explains how the realizers of the property perform the specified tasks. (Kim, 2002, p. 642)

He cannot have it both ways, however. If step i really involves a functional *definition*, then the resulting theory is not an identity theory after all but a version of functionalism. If *P* is functionally *defined*, then it is metaphysically necessary that *x* has *P* iff *x* has a certain causal role *R*; that is, *x* has *P* in some possible world iff *x* has *R* in that world. But this is just functionalism, with its attendant multiple realizability and epiphenomenalism, and Kim doesn't want this.

Perhaps what Kim really has in mind is not functional *definition*, but functional *reference fixing*. Just as 'the clear stuff in nearby lakes and streams' is not part of the definition of 'water' but only serves to fix its reference, the causal role we associate with 'pain' need not be definitional or otherwise necessary but may enter only into reference fixation. If this is the case, then *x* has *P* in some world iff *x* has *R* in the actual world. On this view, there are brain states in other possible worlds that count as pains in those other worlds because of the causal role the states have *here*, even though they may lack these roles in those worlds. This is no longer functionalism, for there is on this account no essential connection between psychological type and causal role, just as there is no necessary connection between water and being in nearby lakes and streams. But then we are back to claiming a necessary identity between brain state type and psychological type, so we can neither invoke the former to explain the occurrence of the latter nor explain how the former realizes the latter. We can, as Kim's step iii suggests, explain how it is that BS-7547 has the causal role it does, but this is not the same as explaining how BS-7547 realizes the property of being a pain, for on the functional reference-fixing view, having the causal role is not the same property as being a pain.

Kim's attempt to combine the virtues of functionalism and TM fails. If mental properties are functionally defined, then they are multiply realizable and explicable, but epiphenomenal. If causal role figures only into reference fixation, then mental properties are neither multiply realizable nor epiphenomenal, but not explicable either.

Computationalist functionalism of the sort I've been describing here—CF—is the most explanatory theory we've ever had in the foundations of cognitive science, perhaps the only explanatory theory we've ever had here.<sup>22</sup> It manages to be so explanatory by offering a deflationary account of mentality, according to which mentality is an inert byproduct of processes that are, at bottom, brute, mechanical, and unintelligent. By doing so, it articulates that vague sense that mentality is not as real as physics and enables nontrivial explanations of *why* mental properties are ever instantiated as well as *how* mental properties are realized by what is really just a heap of molecules.

## 5. Causal Laws Without Causal Properties

Property epiphenomenalism guarantees that mentality will not turn out to be too real, but how to ensure that it is real enough? According to what Kim calls "Alexander's dictum," to be real just is to have causal powers. My position combines

epiphenomenalism with realism and thus must explain where Alexander's dictum goes wrong. The solution, I think, lies in the role psychological properties play in laws of psychology.

### 5.1. Causal Laws

There is good reason to think that there are causal laws of psychology (Davidsonians notwithstanding). The Stroop effect, the word superiority effect, the McCulloch effect, the various laws of psychophysics: all seem to be as good candidates for causal laws as anything. Take, for example, the word superiority effect: recognition of a word facilitates recognition of letters within the word. This is robustly counterfactual-supporting, which makes at least a *prima facie* case for its being a law. If so, it is presumably a causal law; the facilitation relation is to all appearances a causal relation.

Notice, however, that this law, like the typical psychological law, ranges over states or events, rather than properties. Hence, this law's being a causal law requires only that event epiphenomenalism be false, not that property epiphenomenalism be false. (The existence of such laws thus provides an argument against event epiphenomenalism, one that does not rest on intuition or common sense.) For the word superiority effect to constitute a causal law, all that is required is that (a) it *be a law* that recognition of a word facilitates recognition of letters within the word, and (b) word recognition *causes* a decreased reaction time for letter recognition. Property epiphenomenalism threatens neither (a) nor (b). As with many other natural laws, the laws of psychology assert causal relationships among mental *events* (or between mental events and perceptual stimuli or motor output), not among mental *properties*.

### 5.2. The Reality of Psychological Properties

But perhaps the very notion of psychological law—let alone causal psychological law—is suspect if CF is true. One reason for such suspicion has to do with the multiple realizability of mental properties to which CF is committed.<sup>23</sup> The functionalist claim that mental properties are multiply realizable is just the claim that they constitute a wildly disjunctive, heterogeneous class. But this means that there cannot be laws ranging over them; psychological properties would have to be so disjunctive as to be unprojectible, hence not kinds, hence not the stuff over which laws can range.<sup>24</sup> The problem with such an argument is that it assumes from the start that physical kinds are the only kinds there are, the explicit rejection of which is the cornerstone of CF. CF can allow that psychological properties are disjunctive and unprojectible *from the standpoint of physics*, but this is perfectly compatible with there being some other set of laws with respect to which they *are* projectible, some other discipline from the standpoint of which they are sufficiently homogeneous and well-behaved to constitute natural kinds.<sup>25</sup> The kinds of physics are not all the kinds there are, and the laws of physics are not all the laws there are.

The situation, as CF sees it, is this: if it is a law that *As* cause *Bs* and a law that *Cs* cause *Ds*, then it will be true that *As* or *Cs* cause *Bs* or *Ds*, even though this will not be

a law if, e.g., the properties 'A or C' and 'B or D' are sufficiently heterogeneous (Fodor, 1974). But some of these disjunctive properties, though heterogeneous and unprojectible from the standpoint of physics, will be homogeneous and projectible from the standpoint of psychology. It is because they are entailed by true causal laws of physics that they count as *causal* laws, and it is because they are projectible from the standpoint of psychology that they count as causal *laws*. Physics supplies the causality; psychology supplies the lawlikeness.

There is a (mild) sense in which psychological laws are explanatorily impotent: they don't explain any individual events that weren't already explainable. Increasing the set of explained events, however, is not the only way to have explanatory potency. The explanatory contribution made by psychological laws is that of unifying otherwise disparate because physically heterogeneous phenomena. This is not a merely heuristic convenience for finite minds; a Laplacian agent who had no concept of psychological kinds would not be capable of seeing the word superiority effect as a lawlike phenomenon. To the extent that psychological laws and properties and concepts allow us to see it thus, they increase our understanding. Psychological laws thus offer additional explanation(s) without having to increase the number of particular events that get explained. The particular events were already explainable at the physical level, but the instantiation relations, the realization relations, and the higher-level projectibility of a class of physically diverse phenomena all require higher-level laws for their explanation. This is the explanatory contribution of psychological laws.

Psychological properties, though epiphenomenal, are thus not mere Cambridge properties. Though they make no *causal* difference to the world, they do make a difference: they produce no new effects, but they do produce new laws and thereby new explanations. In this sense, CF embraces a stronger realism about psychological properties than even TM, for TM cannot claim any novel contribution on behalf of psychological properties. Alexander's dictum seems plausible only if we assume that something can make a difference only by making a causal difference. This assumption is false. Perhaps a causal difference is the only kind of difference concrete particulars can make, but it is not the only kind of difference properties can make. Instead, we should hold that to be real is to make *some difference or other*, not necessarily to make a causal difference. Psychological properties do make a difference and thus need not be causally efficacious to be real.

Maintaining a strong distinction between psychological laws and physical or neuroscientific laws allows us to distinguish between neural causation and mental causation without any need for efficacious higher-level properties. This is important. The very notion of innateness, for example, rests on a distinction between those changes that are due to growth and those changes that are due to learning, i.e., between biological changes and psychological changes.<sup>26</sup> One way to account for this distinction would be to causally implicate mental properties in some changes and biological properties in the others.<sup>27</sup> Another way, however, is simply to distinguish between those changes that are and those that are not explained by psychological laws. The long-lasting changes that are explained only by biological and lower-level

laws count as innate; the ones that are explained by psychological laws as well count as learned. Similarly, some token brain event causes me to say ‘yes’ and at the same time causes a change in the EEG reading. The former is an instance of mental causation because it is explained by psychological law, the latter of physical causation because it is not explained by psychological law.

### 5.3. *Nonbasic Laws*

This general account will not seem plausible unless we recognize nonbasic laws—laws that are distinct from, but hold as the result of, basic laws of physics. Such laws have been a central feature of NRP since its early formulations (see Fodor, 1974). Absent nonbasic laws, it would be tempting to think that property epiphenomenalism implies event epiphenomenalism, that if it is a law that *Ps* cause *Qs*, then *Ps* must do so *in virtue of* being *Ps*. This, however, is only true of basic laws. While basic laws have the form ‘*Ps* cause *Qs* (and they do so in virtue of being *Ps*)’, nonbasic laws have the form ‘*Ps* cause *Qs* (and they do so in virtue of being *Rs*)’. Of course, basic laws are laws that cannot be explained by other laws; as such, we ought to posit as few of them as possible. If psychological states really had their causal properties in virtue of being the kind of psychological states they are, psychological laws would be basic and hence inexplicable. Proponents of the unity of science and the disunity of science alike want psychological laws to turn out to be nonbasic. Property epiphenomenalism is the only way to ensure this.<sup>28</sup>

The notion of nonbasic laws shows how it is that *Xs* can have causal efficacy even if *X*-ness doesn’t, and this shows how it is possible for event epiphenomenalism to be false even if property epiphenomenalism is true. At the same time, it renders property epiphenomenalism an intuitively plausible view concerning a wide range of properties. Keeping a close eye on the distinction between *Xs* (the states) and *X*-ness (the property), we start to see epiphenomenalism almost everywhere we look. Mountains have climatic effects: rain clouds have a hard time getting over the mountains intact, so the climate on the leeward side of a range is more arid than on the windward side. But mountains don’t have these effects in virtue of their *mountainhood*; they have them in virtue of their height and solidity. A decent fake would do just as well, as would a brick wall of the right size, for that matter. Chairs have causal powers, but again, in virtue of size and solidity, not in virtue of their chairness.

Not only do these verdicts seem intuitively correct, they also provide reassurance to those of us who hope that the causal laws of the special sciences will be explicable. Unless there is something *other than mountainhood* in virtue of which mountains have the causal powers they do, we will not be able to explain why mountains have those causal powers. The properties in virtue of which things have their causal consequences are low-level physical properties whose causal powers cannot be explained in terms of the causal powers of something else. It would be both very surprising and very disappointing if chairness, doghood, or being the belief that

it's raining turned out to be basic properties, if there simply were no explanation for how chairs, dogs, and beliefs that it's raining had the causal powers they do.

Mountainhood *might* be efficacious even though mountains have their causal powers in virtue of something other than mountainhood. Since I am not endorsing any exclusion principles, I am not committed to and have not argued for the claim that *all* nonbasic properties are epiphenomenal, only that functional properties are. However, I see no pressing reason to think that there are nonbasic causal laws that range over causally efficacious properties, and I think the present discussion does much to remove the veneer of absurdity from the claim that only basic properties are efficacious. Since property epiphenomenalism is not refuted by the standard arguments against event epiphenomenalism, and since only *basic* causal laws are required to have the form '*Ps* cause *Qs* (and they do so in virtue of being *Ps*)' the efficacy of *X*-ness is nothing like a precondition on the efficacy of *Xs*.

## 6. Conclusion

The present view requires a delicate balance according to which mentality is real enough to underwrite the role it plays in empirical psychology, yet not so real as to be beyond the reaches of naturalistic explanation. The best way to achieve this balance, I think, is to endorse property epiphenomenalism while denying event epiphenomenalism. As long as mental states figure into causal psychological laws, that seems to be reality enough. But as long as all the effects they have in virtue of something else, they aren't so real as to pose insoluble mysteries. Epiphenomenalism is therefore not an embarrassment for nonreductive physicalism, but actually a virtue of the theory.

## Acknowledgments

Thanks for helpful comments on earlier drafts of this paper or discussions of the material to Chris Hill, Tom Senor, and two anonymous referees for this journal. Some of the main ideas of this paper were discussed in a presentation to the philosophy club at Arkansas State University quite a while back; thanks to the audience in attendance then, especially Don Merrell, Dave Truncellito, and Ron Endicott.

## Notes

- [1] This standard version of NRP proceeds in terms of (static) *states* rather than (dynamic) *events*. This is more terminological than substantive, and what NRP says about states here applies equally to events. In general, I will not mark the distinction between states and events, since nothing will hinge on it for the present purposes. Of primary concern is the distinction between states and events on the one hand, and properties on the other.

- [2] Obviously there is a lot here that is controversial. For details of how one might provide a realist or at least quasi-realist account of proper interpretation, see Cummins (1989), Dennett (1981), and Pylyshyn (1984). For a skeptical view of all this, see Searle (1992).
- [3] In asserting token identity without type identity, CF is committed to denying a property instantiation theory of events (e.g., Kim, 1976) and must instead view them as concrete particulars or the like. This is another feature of CF I won't defend here, but its inclusion is necessary to make sense of the theory itself as well as the possibility of property epiphenomenalism without event epiphenomenalism.
- [4] This is not intended as an affirmation of an abundant over a sparse conception of properties. 'Property' will refer to properties abundantly construed; while 'kind' will refer to a particular type of sparsely construed property.
- [5] Opponents of NRP (e.g., Bickle, 1998; Kim, 1998) sometimes label NRPists as property dualists, perhaps for rhetorical effect. Such a label, however, is doubly misleading. First, NRPists insist on more than one type of *kind*, not, like Jackson (1982), on more than one type of *property*. ('Physical', as it appears in 'physical property', means roughly 'material', while 'physical', as it appears in 'physical kind', means roughly 'having to do with physics'.) In addition, most NRPists insist on far more than just two kinds of kinds, contrary to what the term 'dualism' would suggest. There are physical kinds (i.e., not just material properties, but kinds projected by the laws of physics), but also chemical kinds, biological kinds, etc., supposing these kinds are multiply realizable. Psychology is thus in the same boat as biology here; the NRPist is better thought of as a *kind pluralist* than a *property dualist*.
- [6] Although I will occasionally use pain and the like as convenient examples, CF is first and foremost a theory of cognition and the relation of cognitive psychology to neuroscience and physics. If computationalism can be extended to account for qualia and affect as well, then it can perhaps provide a foundation for all of psychology, not just that part that deals with cognition. Many of computationalist functionalism's more ardent defenders are doubtful of this, however, and the plausibility of CF should not be confused with that of the more ambitious theory. The issue here is not whether CF can account for qualia and the like; worries about epiphenomenalism were supposed to involve a new problem for NRP, not an old one.
- [7] Although he seems to endorse Kim's causal exclusion arguments against NRP, Bickle (1998) explicitly insists that psychological types are multiply realizable. If so, however, he must deny that psychological kinds are identical with neuroscientific kinds, and these causal exclusion arguments would seem to undermine his own reductionist view. The fact that Kim and Bickle both call themselves reductionists should not lead us to think that they are on the same team. It is Kim-style reduction I am opposing here.
- [8] Something like this occurs in philosophy. Most of Kim's arguments for his reductive physicalism are *a priori*, although such a theory entails what Hill (1992) calls the "correlation thesis": that mental state types correlate with neural state types. The latter, however, is an empirical thesis in cognitive neuroscience—as Hill recognizes but Kim seemingly does not—and one for which an *a priori* argument would be inappropriate.
- [9] One could articulate property epiphenomenalism in terms of tropes (e.g., Robb, 1997), but I am looking for a theory-neutral characterization.
- [10] I will be taking this *in virtue of* relation as primitive. It is tempting, though I think inappropriate, to cash it out in counterfactual terms. First, to do so begs the question against causal exclusion principles, since counterfactual dependence is not exclusive. I want to pursue a solution to the problem of epiphenomenalism that does not depend on any controversial theses about causation or causal relevance, on either counterfactual theories or causal exclusion principles. Second, it is unclear how such an account would proceed. Notoriously, causes are often counterfactually dependent on their effects as well as vice versa. Also notorious is the ease with which we can conjure up scenarios where an event counterfactually depends on something that is intuitively causally irrelevant. An evil

neuroscientist has determined to fry my visual cortex if neuron number 45883 (N-45883), which is somewhere in my peripheral nervous system, doesn't fire at  $t$ , thus precluding my having any visual experiences at  $t + \delta$ . Nonetheless, N-45883's firing isn't causally relevant in any interesting sense to my having a visual experience at  $t + \delta$ . Certainly, my visual experience does not occur *in virtue of* N-45883's firing. Such prosaic considerations do not of course show that a counterfactual theory could not work; they are merely intended to justify taking the *in virtue of* relation as primitive. The intuitive appeal of the counterexamples shows that the *in virtue of* relation is clear enough for ordinary purposes without analysis.

- [11] At the risk of being pedantic: Although it is conventional to imagine that being in pain might be type-identical with having one's C-fibers firing and that this might be of use to the identity theorist, this is potentially very confusing, given that there really are such things as C-fibers. One cannot always tell whether a given author is really talking about real C-fibers (i.e., the unmyelinated nociceptive neurons with dendritic termina in the skin and which synapse onto the dorsal horn of the spinal chord) or is simply using the term to allude to a convenient fiction. Real C-fibers won't do the trick. Being in the *peripheral* nervous system, C-fiber firing is known to be neither (nomologically) necessary nor sufficient for pain, and thus cannot be identical with it. Pain presumably occurs in the *central* nervous system, in particular, in the brain. Finally, it is far from obvious that real C-fibers are not already functionally defined, as the use of the term 'nociceptive' above suggests.
- [12] I do *not* mean to be raising the question of whether qualia in general are epiphenomenal. There is every reason to think that the same sort of argument would work for cognitive states, though it would be more cumbersome, and I won't pursue the details here.
- [13] Obviously it is an empirical question just what kind of physical property realizes intensity of pains and an empirical question how ubiquitous the many-to-one mappings discussed here are going to be. Consequently, how *much* property epiphenomenalism the exclusion arguments would entail is an empirical question. Given the empirical facts that we do know, however, it seems pretty clear that causal exclusion arguments will yield at least *some* unintended epiphenomenalism. That is, there will be some property of some mental state, which strikes us as being efficacious, but which is realized by some physical property such that not all and only the physical states having that physical property realize mental states having that mental property.
- [14] Kim at least is very explicit in his 1992 article about the unreality of disjunctive kinds. He thinks that if a property is multiply realizable, then it is not a genuine kind after all. As far as I can see, the empirical facts about individual differences in hemispheric specialization and localization of function more generally lead such a view to eliminativism about language comprehension as well as many other psychological phenomena.
- [15] Block (1990a) comes fairly close to making this point, though no one to my knowledge has been explicit about it.
- [16] This raises an interesting and important problem regarding the relation between psychological and functional properties. On the face of it, if the functional properties are metaphysically prior to the psychological properties, as most functionalists seem to think, then the latter cannot be identical with the former, as most functionalists also seem to think. However, I will not try to decide here how to resolve this apparent conflict.
- [17] I do not deny that there may be explanatory contexts where it is appropriate to say that it held the door open because it was a doorstop; the present point is not one about context sensitive matters of explanation but rather about metaphysical priority, which presumably is not context dependent.
- [18] I have been working under the simplifying assumption that *each* of the effects of some state is essential to its functional individuation, whereas it is more plausibly the case that only the *whole complex* of effects is essential or that the instantiation of a functional property only probabilifies but does not entail most individual token effects. For instance, it may be

necessary that if one is in pain, then one is disposed to cry out, but my being in pain at  $t$  does not entail my crying out at  $t$ . Nonetheless, discharging the simplifying assumption would make little substantive difference, for there are sure to be *some* effects that are individually necessary, and epiphenomenalism is guaranteed with respect to these. Thus, to continue with the present concrete example, Loewer's account would only make room for causal efficacy with respect to the *incidental* effects of mental states. While its being a pain did not causally contribute to the avoidance response (which, let us suppose, is an essential component of the pain role), it did causally contribute to the deflection of some electrons (since that is not essential).

- [19] Certainly TM can claim that psychology reduces to physics rather than the other way around, if, for instance, all the laws of psychology are derivable from all the laws of physics (conjoined with the appropriate bridge laws or what have you), but some laws of physics are not thus derivable from the laws of psychology. However, the import of identifying psychological *kinds* with physical *kinds* is that the laws of psychology thus turn out to be notational variants on laws of physics; that is, they just are laws of physics. There is presumably a set of fundamental, quasi-axiomatic laws of physics from which these other laws of physics can be derived, and TM has no trouble explaining how psychological laws can be less fundamental than these fundamental laws. What it cannot obviously account for is how psychological laws can be—as they presumably are—less fundamental than the *derived* laws of physics.
- [20] There will only be three if we identify  $M$  with the disjunction  $P1 \vee P2$ , and this is an option open to CF, though I will not pursue it here.
- [21] One additional constraint that is frequently imposed is that the semantic interpretation actually add something: that it discover “real patterns” (Dennett, 1981, 1991) not discoverable without it or that it produce new nomic generalizations (Pylyshyn, 1984). This will typically require that the system in question be sufficiently complex; the constraint is intended to rule out attributions of mentality to thermostats and the like.
- [22] The few historical theories that have even come close to explaining how physical phenomena give rise to mental phenomena have been limited in scope and have been mostly crypto-computationalist theories anyhow (in fact, mostly crypto-connectionist theories, like Hebb's 1949 account of associative learning). One noteworthy exception, ironically, is Descartes' (1649/1985) hydraulic theory of, among other things, how rehearsal facilitates mnemonic recall. It goes without saying that Descartes did not purport to have a physicalistic account of the whole cognitive mind.
- [23] Another is the Davidsonian claim, alluded to above, that laws need to be strict, and psychological generalizations are never strict. I lack the space to address this here, except to say that I see no reason at all to claim that laws must *ipso facto* be strict. Everything I say here will, however, be compatible with the thesis that only strict laws are capable of securing causal efficacy on a covering law model of causation (though I doubt that even this is true, given that the fundamental laws of physics are so unlikely to turn out to be strict).
- [24] Kim seems to have this sort of argument in mind. See his discussions of projectibility in Kim (1992, 1996, 1998).
- [25] This seems to be roughly the view of Fodor's older 1974 special sciences paper, though he does not rely on it in his newer 1997 special sciences paper.
- [26] The notion of nativism in psychology is thus quite different from the notion of nativism in biology, where innateness may be roughly a matter of independence from environment (Ariew, 1996). In psychology, it is a matter of what kind of environmental causes are operative.
- [27] As Eric Funkhouser pointed out to me in conversation.
- [28] We often say things like ‘ $P$ s cause  $Q$ s (and they do so in virtue of being  $R$ s)’ without intending to suggest that ‘ $R$ s cause  $Q$ s’ is itself a basic law. I suggested earlier, for instance, that Mt. Everest causes people to try to climb it, not in virtue of being the tallest mountain,

but in virtue of *being believed to be* the tallest mountain. Clearly, such talk is not intended to be the final word concerning which properties are ultimately causally efficacious but only to assert that some property, e.g., being the tallest mountain, is *not* efficacious.

## References

- Ariew, A. (1996). Innateness and canalization. *Proceedings of the Biennial Meetings of the Philosophy of Science Association*, 3, S19–S29.
- Bickle, J. (1998). *Psychoneural reduction: The new wave*. Cambridge, MA: MIT Press.
- Block, N. (1990a). Can the mind change the world? In G. Boolos (Ed.), *Meaning and method* (pp. 137–170). Cambridge, England: Cambridge University Press.
- Block, N. (1990b). The computer model of the mind. In E. E. Smith & D. N. Osherson (Eds.), *An invitation to cognitive science: Thinking* (Vol. 3, pp. 147–289). Cambridge, MA: MIT Press.
- Block, N. (1995). The mind as the software of the brain. In E. E. Smith & D. N. Osherson (Eds.), *An invitation to cognitive science: Thinking* (2nd ed., Vol. 3, pp. 377–425). Cambridge, MA: MIT Press.
- Churchland, P. M. (1988). *Matter and consciousness*. Cambridge, MA: MIT Press.
- Cummins, R. C. (1989). *Meaning and mental representation*. Cambridge, MA: MIT Press.
- Cummins, R. C. (1992). Conceptual role semantics and the explanatory role of content. *Philosophical Studies*, 65, 103–127.
- Cummins, R. C. (1996). *Representations, targets, and attitudes*. Cambridge, MA: MIT Press.
- Dennett, D. C. (1981). True believers: The intentional strategy and why it works. In *The intentional stance* (pp. 13–35). Cambridge, MA: MIT Press.
- Dennett, D. C. (1991). Real patterns. *Journal of Philosophy*, 88, 27–51.
- Descartes, R. (1985). *The passions of the soul*. In J. Cottingham, R. Stoothoff, & D. Murdoch (Trans.), *The philosophical writings of Descartes* (Vol. 1, pp. 325–404). Cambridge, England: Cambridge University Press. (Original work published 1649)
- Fodor, J. A. (1974). Special sciences. *Synthese*, 28, 97–115.
- Fodor, J. A. (1975). *The language of thought*. New York: Crowell.
- Fodor, J. A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1989). Making mind matter more. *Philosophical Topics*, 17, 59–79.
- Fodor, J. A. (1997). Special sciences: Still autonomous after all these years. In J. E. Tomberlin (Ed.), *Philosophical perspectives: Mind, causation, and world* (Vol. 11, pp. 149–163). Boston: Blackwell.
- Harnish, R. M. (2002). *Minds, brains, computers*. Malden, MA: Blackwell.
- Haugeland, J. (1981). Semantic engines: An introduction to mind design. In J. Haugeland (Ed.), *Mind design* (pp. 1–34). Cambridge, MA: MIT Press.
- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
- Hill, C. S. (1992). *Sensations: A defense of type materialism*. Cambridge, England: Cambridge University Press.
- Horgan, T. (1989). Mental quasation. *Philosophical Perspectives*, 3, 47–76.
- Horgan, T. (2001). Causal compatibilism and the exclusion problem. *Theoria*, 16, 95–116.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–136.
- Kazez, J. (1994). Computationalism and the causal role of content. *Philosophical Studies*, 75, 231–260.
- Kim, J. (1976). Events as property exemplifications. In M. Brand & D. Walton (Eds.), *Action theory* (pp. 159–177). Dordrecht, Netherlands: Reidel.
- Kim, J. (1989). The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association*, 63, 31–47.
- Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research*, 52, 1–26.
- Kim, J. (1993). *Supervenience and mind*. Cambridge, England: Cambridge University Press.

- Kim, J. (1996). *Philosophy of mind*. Boulder, CO: Westview.
- Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: MIT Press.
- Kim, J. (2002). Précis of *Mind in a physical world*. *Philosophy and Phenomenological Research*, 65, 640–643.
- LePore, E., & Loewer, B. (1987). Mind matters. *Journal of Philosophy*, 84, 630–642.
- LePore, E., & Loewer, B. (1989). More on making mind matter more. *Philosophical Topics*, 17, 175–191.
- Lewis, D. K. (1969). Review of W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion*. *Journal of Philosophy*, 66, 23–25.
- Loewer, B. (2002). Comments on Jaegwon Kim's *Mind and the [sic] physical world*. *Philosophy and Phenomenological Research*, 65, 655–662.
- Ludwig, K. A. (1994). Causal relevance and thought content. *Philosophical Quarterly*, 44, 334–353.
- Marr, D. (1982). *Vision*. London: Allen and Unwin.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Polger, T. W. (2004). *Natural minds*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
- Robb, D. (1997). The properties of mental causation. *Philosophical Quarterly*, 47, 178–194.
- Searle, J. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Copyright of *Philosophical Psychology* is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.