

Evolution, Epiphenomenalism, Reductionism

ALVIN PLANTINGA
University of Notre Dame

I Semantic Epiphenomenalism

A common contemporary claim is the conjunction of metaphysical naturalism—the idea, roughly, that there is no such person as God or anything at all like God—with the view that our cognitive faculties have come to be by way of the processes to which contemporary evolutionary theory direct our attention. Call this view ‘N&E’. I’ve argued elsewhere¹ that this view is incoherent or self-defeating in that (1) anyone who accepts it has a defeater for R, the proposition that her cognitive faculties are reliable, which then gives her (2) a defeater for any proposition she believes, including, of course, N&E itself. The argument for (1), in turn, depends essentially on the proposition that (3) $P(R/N\&E)$ is low or inscrutable. To support (3), I divided N&E into mutually exclusive and jointly exhaustive subcases, arguing that in each subcase S_i , $P(R/N\&E\&S_i)$ is low or inscrutable. I won’t repeat this argument here, but I do want to focus on a certain essential aspect of the argument for (3).

But first we must note that one who accepts metaphysical naturalism will likely be a materialist or a physicalist with respect to human beings: materialism is almost universally thought to be *de rigueur* for naturalists. So let’s at least temporarily assimilate materialism with respect to human beings to naturalism. Now suppose materialism is true: given that there *are* such things as beliefs, what sort of thing will a belief *be*?² Since we are assuming materialism, it will presumably have to be a material process or event—perhaps a long-standing neural event of some kind in the nervous

¹ *Warrant and Proper Function* (New York: Oxford University Press, 1993), chapter 12; *Warranted Christian Belief* (New York: Oxford University Press, 2000) pp. 227ff.

² A materialist might hold that human beings sometimes display the property of believing *p*, for some proposition *p*, while denying that there are any such things as *beliefs*; belief talk would then be paraphrased into talk about the property of believing. For what follows this difference will make no difference.

system.³ Perhaps it will be a neural event involving some of the approximately 10^{11} neurons most of us enjoy; perhaps it will involve these neurons being connected in a certain fashion, firing in a certain sort of pattern, responding in a certain way to differential input from other neural processes, and so on. So considered, beliefs will of course enter the causal process that leads to behavior.

So considered, however, beliefs will have two quite different sorts of properties. On the one hand, such an event will have neurophysiological properties—number of neurons and neural connections involved, strength and rate of firing at different times and in various parts of the event, rate of change of strength and rate of firing in response to differential input, and the like. But if this neural event is really a *belief*—perhaps the belief that Proust was a more subtle writer than Dickens—it must have another kind of property as well. It will have to be the belief that p for some proposition p —in this case the proposition that Proust was more subtle than Dickens. Some proposition p (ignoring vagueness) will be uniquely associated with that neural event as its *content*: the proposition p such that the belief in question is the belief that p . And by virtue of having content, this belief-event will also have intentionality or aboutness. That is because the proposition that is the content of the belief will be directed on some subject matter (for example, Proust and Dickens), and predicate something of that subject matter (for example: that the former is more subtle than the latter); and the belief inherits this intentionality. So if materialism is true, then beliefs will be neural structures that display these two kinds of properties.

Now as we all know, it is far from clear how a neural structure could acquire a content. How does a neural event somehow reach out and seize a given object—perhaps the Taj Mahal—so that it is *about* that object? How does it somehow get assigned a certain proposition as its content? It is hard to think of any scenarios that are as much as decently plausible.⁴ Still, it must happen somehow, at least if materialism is to accommodate the idea that there really are such things as beliefs. So let's assume for the moment that such a structure can acquire a content. The next problem is that it seems extremely difficult to see how that structure can enter the causal chain leading to behavior *by virtue of its content*. It is easy enough to see how it can enter the causal chain by virtue of its neurophysiological properties; current science gives us a reasonably full and detailed account of the process whereby volleys of impulses propagated along the efferent nerves cause muscle contraction, motor output, and thus behavior. But how does the *content* of this belief, the

³ Note that if property dualism is true, these events could be mental events in that they (or the objects involved in them) display mental properties; they are still material events, however, in that only material substances will be involved in them.

⁴ In *Explaining Behavior* (Cambridge, MA: The MIT Press, 1988) Fred Dretske makes a valiant but, I think, wholly unsuccessful attempt to provide such a scenario.

fact that it has this particular proposition as content—how is *that* relevant to the causal powers of the belief? As Fred Dretske puts it,

We can, following Davidson, say that reasons *are* causes, but the problem is to understand how their being reasons contributes to, or helps explain, their effects on motor output. The fact that they have a content, the fact that they have a *semantic* character, must be relevant to the kind of effects they produce. If brain structures possessing meaning affect motor output in the way the soprano's acoustic productions affect glass, then the meaning of these neural structures is causally inert. Even if it is there, it doesn't *do* anything. If having a mind is having *this* kind of meaning in the head, one may as well not have a mind.⁵

Dretske goes on to note that this problem is clearly an analogue of the one often thought to afflict Cartesian dualism: the problem of explaining how immaterial events or structures (*beliefs*, e.g., which on a dualistic view will be modes of an immaterial substance) can causally affect the hard, heavy, massive material world. For the materialist, the analogous problem is that of specifying how a neural structure's propositional *content* can causally influence behavior—more exactly, how it is that it causally influences behavior by virtue of having the content it does have. When the soprano hits high C and shatters the champagne glass, it is not by virtue of the content or meaning of the line she sings that the glass is shattered; it is just by virtue of the physical properties of the event in question. Even if the words she sings had a wholly different and indeed contrary content—instead of “Oh how I love him!” they mean “Oh how I hate him!”—the glass would have shattered in the same way. The content seems to be causally impotent, causally irrelevant. And how can the materialist avoid the same conclusion in the case of the content enjoyed by those neural structures that are beliefs? Suppose a given neural structure both plays a causal role in the production of motor input and also has a content: that Proust is more subtle than Dickens. It seems that the causal role it plays depends just on those neurophysiological properties it possesses: would it not produce the same physical effect, the same effect on nerves and muscles and glands, even if it had a quite different content, indeed, even if its content were the proposition that Dickens is more subtle than Proust?⁶ Suppose we use the term ‘semantic epiphenomenalism’ (‘SE’, for short) as a name for the proposition that the content of belief is causally irrelevant to behavior. From the perspective of materialism, SE is a haunting specter and cause for anguish. Thus Jerry Fodor:

I'm not really convinced that it matters very much whether the mental is physical; still less that it matters very much whether we can prove that it is. Whereas, if it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my

⁵ *Explaining Behavior*, p. 80.

⁶ Of course if content supervenes on neurophysiological properties (together, perhaps, with certain environmental properties) then it won't be possible that this event have the same neurophysiological properties but a different content: see below, p. 619.

scratching, and my believing is causally responsible for my saying ... if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world.⁷

To return to the argument for (3), it looks initially as if $P(SE/N\&E)$ ($N\&E$ construed as including materialism) is high. But it also seems that $P(R/N\&E\&SE)$ is low (or perhaps inscrutable). We can see this as follows. Given a certain array of behavior-producing structures, natural selection can modify these structures in the direction of greater fitness or adaptiveness, according a reproductive edge to those organisms sporting mutations (mutations involving these structures) that permit or enhance reproductive success. Natural selection can eliminate certain structures (those that come at the cost of reproductive fitness) and encourage others (those that enhance reproductive fitness). Now if false belief caused maladaptive action, natural selection could presumably modify belief-producing structures in the direction of greater reliability—a greater proportion of true as opposed to false belief. But if content does not enter the causal chain that leads to behavior, then of course it will not be the case that a belief causes maladaptive behavior by virtue of its being false, and it will not be the case that a true belief causes the behavior it does by virtue of being true. And then it is hard to see how natural selection can promote or enhance or reward true belief and penalize false belief.

Now the *reliability* of a cognitive faculty or process requires the truth of at least a substantial proportion of the beliefs it produces.⁸ But suppose semantic epiphenomenalism is true. Then by modifying behavior (more exactly, the structures that produce behavior) in the direction of greater fitness, natural selection would not necessarily be modifying belief (the structures that produce belief) in the direction of greater reliability, greater proportion of true belief. Indeed, it would be an enormous cosmic coincidence, and enormous piece of not-to-be-expected serendipity—if modification of behavior in the direction of fitness also modified belief-producing mechanisms in the direction of greater reliability. How then should we think of the probabilities involved? We could proceed in more than one way. First, we might ask about the probability of a *given* belief B 's being true, on $N\&E$ & SE . About all we can claim to know here is that B is probably fitness-enhancing by virtue of its neurophysiological properties—or better, that it has been produced by structures that are fitness-enhancing. But this tells us nothing at all about the likelihood of the *truth* of B . Presumably, therefore, we should think that this belief is about as likely to be true as to be false; this probability is in the neighborhood of .5. But if so,

⁷ "Making Mind Matter More," *Philosophical Topics*, 1989, p. 77.

⁸ Strictly speaking, what reliability really requires is that the ratio of true beliefs to false, integrated over the appropriate nearby possible worlds, be high. Accordingly, it is *possible* that our cognitive faculties are reliable, but produce mostly false beliefs in alpha, the actual world—possible, but very unlikely.

then it will be monumentally improbable that a person's whole structure of beliefs displays the preponderance of true belief over false that reliability requires. If I have 1000 beliefs, for example, the probability, on chance alone, that 3/4 or more of these beliefs are true (certainly a reasonable requirement for reliability) will be less than 10^{-58} .⁹ And even if I am running a modest epistemic establishment of only 100 beliefs, the probability that 3/4 of them are true, on chance alone, is something like 1 out of a million. Looking at the matter from this perspective, we should judge that $P(R/N\&E\&SE)$ is very low indeed.

On the other hand, we might think that *one* relevant possibility is that content gets attached to belief in this random fashion, but another is that it is somehow mainly *true* content, *true* propositions that get attached to beliefs with adaptive neurophysiological properties, and still another is that it is mainly *false* content that gets so attached. If you think these possibilities are about equiprobable, then the thing to think is that $P(R/N\&E\&SE)$ is in the neighborhood of .33. But (perhaps less likely) you might also think that we have no sensible way at all of estimating these probabilities, in which case you should take $P(R/N\&E\&SE)$ to be inscrutable. So it looks initially as if $P(SE/N\&E)$ is high and $P(R/N\&E\&SE)$ is low (or possibly inscrutable).

Now according to the probability calculus,

$$(4) P(R/N\&E) = P(R/N\&E\&SE) \times P(SE/N\&E) + P(R/N\&E\&-SE) \times P(-SE/N\&E),$$

i.e., the probability of R on N&E is the weighted average of the probabilities of R on N&E&SE and N&E&-SE—weighted by the probabilities of SE and -SE on N&E. But inspection of (4) shows that if $P(SE/N\&E)$ is high and $P(R/N\&E\&SE)$ is low, then $P(R/N\&E)$ is also low. For example, if $P(SE/N\&E)$ is .9 and $P(R/N\&E\&SE)$ is .2, then even if $P(R/N\&E\&-SE)$ is 1, $P(R/N\&E)$ is .28. Of course it is ludicrous to assign precise values to these probabilities; still our estimates of them can be guided by (4).

So it looks initially as if $P(SE/N\&E)$ is high; given that $P(R/N\&E\&SE)$ is low, this yields the result that $P(R/N\&E)$ is low, just as (3) has it. This is how it looks *initially*; appearances are sometimes misleading, however: is there some way in which N&E&Materialism can be reconciled with the idea that the content of belief plays a role in the causation of behavior? Is there any way in which the materialist can ward off semantic epiphenomenalism?

⁹ My thanks to Paul Zwier, who performed the calculation.

II Kim Reductionism

A. *What is it?*

Perhaps. In fact there are several ways in which a materialist might try to avoid, elude, or otherwise sidestep SE. With Jerry Fodor, for example, he might argue that there are counterfactuals relating behavior with belief; or with Tyler Burge and Lynne Rudder Baker he might claim that there are laws connecting the two, which suffices for escaping SE; or with Stephen Yablo he might see the relevant phenomena related as determinates and determinables, and hold that this is sufficient for dodging SE. Here I want to examine one particularly promising attempt to avoid SE, the reductionism offered by Jaegwon Kim. The dust jacket of Kim's *Mind in a Physical World*, the most recent book length presentation of his views on that topic, displays the following quotation from Brian McLaughlin: "This elegant work is filled with philosophical wisdom and insight.... It articulates beautifully what the mind-body problem is for us today." This work is indeed elegant and does indeed contain much philosophical wisdom and insight. And while it doesn't articulate the mind-body problem for those of us obstinates who remain dualists, as a materialist account of mind it is about as competent, insightful, and penetrating as anything one can find. Acting on the principle that nobody's perfect, however, I want to raise some problems for Kim's main theses; I'll argue that his reductionism is not successful. And I'll also argue that even if his reductionism *is* successful, it offers no help in resisting the evolutionary argument against naturalism with which we began.

Kim starts by attacking *nonreductive* materialism, according to which (1) physicalism is true: human beings are material objects with no immaterial components or parts, and the physical domain is causally closed, and (2) mental properties supervene on (in Kim's well-known sense) but are not reducible to physical properties. He argues that nonreductive materialism faces grave problems with respect to mental causation—the very problem that induced that apocalyptic outburst of Fodor's. This problem, he says, is that nonreductive materialism implies epiphenomenalism with respect to mental properties: such properties are *causally impotent* in that things and events do not have effects by virtue of having them. I don't have the space to consider his arguments here, but I believe considerations in the neighborhood of his arguments really do pose powerful problems for nonreductive materialism.

After this attack on current orthodoxy Kim turns to his own positive contribution: *reductive* materialism. Here, initially at any rate, the idea is threefold: (1) physicalism is indeed true, (2) the mental supervenes on the physical, and (3) mental properties *can be reduced* to physical properties. Part of the point of this position is that it is supposed to show that and how mental events really do cause behavior, and do so by virtue of their mental properties. Specified to our topic, the idea is to show how it is that my belief

there is a beer in the fridge is a part cause of my body's moving over to the fridge and extracting a beer—and is a cause by virtue of its content, not just its neurophysiological properties.

In chapter 4, pp. 97-103, the basic idea of Kim reduction seems to be that the reducee, a given mental property M, say, is reduced to a physical property P by being shown to be *identical with* P—but then, by Leibniz's Law, M will of course have all the causal properties enjoyed by P. Since the physical property in question is in the typical case causally active, SE is thus parried. But how does the reduction go? As follows. First, the mental property—that of believing that Proust is more subtle than Dickens, for example—is *functionalized*: that is, declared to be a *functional* property: "For functional reduction we construe M as a second-order property defined by its causal role" (98). M is the property that a thing x has just if x has a property that plays a certain causal role, perhaps that of mediating between certain kinds of sensory input and motor output. *Believing that Proust is more subtle than Dickens*, then, is the property of having a property that plays that causal role: it is therefore a second-order property. It is the property of having a first-order property that meets a certain description.

That's the first step; the second step in the reduction is to reduce this functional property to the (physical) property that does in fact meet that description—in this case, that of playing that causal role mentioned above. And the reduction, again, proceeds by *identifying* the former with the latter. In the present case, then, there is the property *believing that Proust is more subtle than Dickens*. This property is the property of having a property that plays a certain causal role C. And *that* property—the property *having a property that plays causal role C*—is identified with whatever physical property P it is that does in fact play this causal role. But identity is transitive: thus the property of believing that all men are mortal is identical with the physical property that plays that causal role C. And the problem with the causal role of this mental property is allegedly solved, for clearly that mental property will have whatever causal powers are enjoyed by the physical property with which it is identical—just as the first person to climb Mt. Rainier has all the causal properties of E. van Trump, the latter being identical with the former.

Sadly enough, however, a problem looms. In fact two problems loom. First, there is the problem of multiple realizability: presumably it is possible that there be creatures—extraterrestrials, let's say—of very different cognitive architecture who nonetheless could form the belief that Proust is more subtle than Dickens. But if so, then it is hard to see how there could be a particular and specific neurological property such that displaying *that* very property is necessary for holding that belief. In fact the problem doesn't require recourse to extraterrestrials. According to the wholly reputable journal *Science*, a

researcher into hydrocephalus reports on someone who had brain material measuring only about 1/45 of that enjoyed by most of us, but who nevertheless carried on a perfectly normal cognitive life:

There's a young student at this university, says Lorber, "who has an IQ of 126, has gained a first-class honors degree in mathematics, and is socially completely normal. And yet the boy has virtually no brain." The student's physician at the university noticed that the youth had a slightly larger than normal head, and so referred him to Lorber, simply out of interest. "When we did a brain scan on him," Lorber recalls, "we saw that instead of the normal 4.5 centimeter thickness of brain tissue between the ventricles and the cortical surface, there was just a thin layer of mantle measuring a millimeter or so. His cranium is filled mainly with cerebrospinal fluid."¹⁰

No doubt this student believes that Proust is more subtle than Dickens; but it is hard to imagine that there is some physical, presumably neurological property P such that this student and a normal human being both display P and such that P just is the property of having the belief that Proust is more subtle than Dickens.

That's the first problem; there is another that is more difficult. The second-order property

- (5) having a property that meets condition C

can hardly be identical with whatever first-order property it is that meets condition C, just as the property

- (6) having Marian's favorite property

i.e.,

- (6) having a property P such that Marian prefers P to any other property

can hardly be identical with *being an early riser*, even if the latter is in fact Marian's favorite property. It is only a contingent fact that *being an early riser* is Marian's favorite property; in some other possible world, where his favorite property is that of lying abed until noon, a person could have a property that was Marian's favorite without having the property *being an early riser*. Hence, by Leibniz's Law, the latter is not identical with the former.¹¹

Kim is certainly aware of this problem (see footnote 11, p. 132), but how does he propose to deal with it? In a rather breathtaking maneuver, he declares that there really *aren't* any second-order properties; there aren't any properties

¹⁰ Roger Lewin "Is Your Brain Really Necessary?" *Science* 210 (1980), p. 1232.

¹¹ In this connection, the argument on the bottom of p. 98 and the top of p. 99 seems wholly fallacious.

that involve quantifying over properties; and hence (apparently) there aren't any functional properties. "By quantifying over properties, we cannot create new properties anymore than by quantifying over individuals we can create new individuals.... So it is less misleading to speak of second-order *descriptions* or *designators* or second-order *concepts*, than second-order properties" (104). There isn't any such thing as the property of being Marian's favorite property or the property of having a property that plays causal role C. But then what about the claim that mental properties are to be identified with physical properties? If, as Kim seems to hold, mental properties are functional properties, then if there aren't any functional properties, there presumably won't be any mental properties either—in which case a reduction of mental properties to physical properties hardly seems possible, let alone necessary or desirable. So how are we to understand Kim reduction?

The following passage seems to be a summary of Kim's reductive strategy:

So where does all this leave us as regards reduction and reductionism? Let M be a mental property ... and let us suppose how we should view the situation when the functional model of reduction is brought to bear on M.... A functional reduction of M requires the functionalization of M; let us assume that this has been done. We also assume that M has multiple physical realizers in different species and structures and can have different realizers in different possible worlds. The reduction consists in identifying M with its realizer P_1 relative to the species or structure under consideration (also relative to the reference world). Thus M is P_1 in species 1, P_2 in species 2, and so on. (110.)

This passage is puzzling along more than one dimension. First, here we seem to be assuming that there *are* mental properties, and that they can be identified (via 'functionalization') with functional properties. But the preceding five pages have been devoted to an argument for the conclusion that there *aren't* any functional properties; and hence, one thinks, also no mental properties, at any rate if, as Kim says, mental properties are functional properties. So how is this reduction to be accomplished? I must confess I don't really see. But second, if I have Kim at all right, the driving intuition, the payoff, the bottom line, however it is supposed to be arrived at, is that *believing that Proust is more subtle than Dickens* is P_1 in S_1 , P_2 in S_2 , and so on: "The reduction consists in identifying M with its realizer P_1 relative to the species or structure under consideration (also relative to the reference world). Thus M is P_1 in species 1, P_2 in species 2, and so on" (110). This too is puzzling in the extreme: if we are talking identity, genuine identity, what is that phrase 'in S_1 ' doing there? That seems a little like saying that in Virginia, George Washington was (i.e., was identical with) the most distinguished American philosopher. That's no doubt true, but it's also misleading; it isn't as if in Rhode Island, e.g., it might be that someone else was identical with the most distinguished American philosopher. It's not

possible that there be a certain property, M , which is identical with P_1 in one species but with P_2 , a different property, in a different species. That can't be, of course, because identity, real identity, isn't relative to species. If M really is identical with P_1 in species S_1 , then M_1 is identical with P_1 *simpliciter*, and the addendum 'in S_1 ' is irrelevant. Further, as we all also know, identity is an equivalence relation: so if M really is identical with P_1 in species S_1 , then M is indeed identical with P_1 ; but then by the same token it is identical with P_2 ; hence P_2 is identical with P_1 , which by hypothesis is false. So we can't take the proposed reduction at face value.

But then how *are* we to take it? Perhaps the following contains the clue: "In this way multiply-realized properties are sundered into their diverse realizers in different species and structures, and in different possible worlds" (111). These multiply-realized alleged properties are *sundered* into their diverse realizers. I believe Kim's view is that there really isn't any such *property* as *believing that Proust is more subtle than Dickens*, a property exemplified by all the creatures of various species and structures who are correctly said to believe that Proust is more subtle than Dickens; what there are instead are mental *concepts* like *believing that Proust is more subtle than Dickens* and mental *designators* like 'believing that Proust is more subtle than Dickens' (104-105). What there are instead of mental properties are species or structure specific physical properties like P_1 , P_2 , and the like. There isn't any such property as *believing that Proust is more subtle than Dickens*; what there is instead is P_1 , which realizes *believing that Proust is more subtle than Dickens* in structure S_1 , P_2 , which realizes it in structure S_2 , and the like. So we might speak here of *H-belief that Proust is more subtle than Dickens*, which would be the physical property that realizes *believing that Proust is more subtle than Dickens* in human beings.¹² So much seems fairly solid, as a construal of Kim; but again, how is the reduction supposed to go? I'll return to that question. But first, we should note that just this much already leads to serious problems; for it seems to commit Kim to the very semantic epiphenomenalism for which his position is supposed to be the remedy.

B. Kim Reductionism: Semantic Epiphenomenalism?

Recall that the main point of Kim reductionism is to avoid what we might call 'mental property epiphenomenalism'. A given neural structure may have both mental and physical properties; and the point is to show that and how the mental properties are causally effective. To turn to semantic epiphenomenalism, the variety of mental property epiphenomenalism presently at issue, consider the belief that all men are mortal: as we have

¹² Ignoring for the moment the likely fact that belief is very likely realized by different properties in different human beings, or even the same human being at different times.

seen, this belief will be a neural structure or event of some kind, with physical (neurophysiological) properties; but it will also have a certain content. It will be related to a certain proposition in the way in which the belief that *p* is related to the proposition that *p*. So here Kim reductionism, if it is successful, will show how it is that *holding a belief with the content all men are mortal* plays a role in causing behavior; alternatively, how it is that *being a belief with the content all men are mortal* plays a role in causing behavior.

Now Kim reductionism originally promised to accomplish this feat by showing how a certain mental property—*believing that all men are mortal*—is identical with a physical property *P*. As we have seen, however, Kim also seems to hold that there isn't any such thing as the mental *property believing that all men are mortal*. What there is instead is the mental *concept*. Now Kim doesn't say much about concepts. If a belief is a neural event of a certain kind, perhaps the same should be said of concepts. Perhaps some concepts are something like *grasps* or *apprehensions* of properties; this is perhaps implied by Kim's suggestion that "In building scientific theories, we hope that the concepts in our best theories pick out, or assent to, the real properties in the world" (105). Apparently, then, there are concepts corresponding to properties—corresponding to individual single properties; call these 'primary concepts'. For example, perhaps there is a concept corresponding to the property *weighs more than a ton*. But there are also other concepts that don't at any rate correspond to a single property: disjunctive concepts, and ones that involve quantifying over properties would be of this sort. Call these 'secondary concepts'. In any event, however, concepts will *apply to* objects; objects will *fall under* concepts. An object falls under a primary concept just if it has the property corresponding to that concept; it falls under a secondary concept just if it has a property corresponding to one of the disjuncts of the concept. Now suppose *believing that all men are mortal* gets multiply-realized; what sort of thing is it that gets realized in this way? The only candidate in the neighborhood would be the concept *believing that all men are mortal*. So what gets multiply-realized is not a mental property, but the mental concept *believing that all men are mortal*. The physical property *H-believing that all men are mortal* *realizes* that concept. It isn't completely clear what this realization relationship is; the relation between *believing that all men are mortal* and *H-believing that all men are mortal* is equally unclear. What does seem fairly clear, however, is that if the latter realizes the former in a given structure *S*, then the conjunction of the latter with the property of displaying *S* *entails* the former. Perhaps there is no such property as *believing that all men are mortal*; nevertheless there is the *concept believing that all men are mortal*, a concept such that anything that falls under it believes that all men are mortal; and presumably anything

that is human and has the property *H-believing that all men are mortal* falls under the concept *believes that all men are mortal*. Otherwise it would be extremely hard to see the relevance of H-belief to belief.

Now what would have to be the case for semantic epiphenomenalism to be *false*? Well, of course it would have to be the case that beliefs are causally active by way of their content. That is to say, it would have to be the case that the property *holding a belief with the content that all men are mortal* is causally efficacious. If semantic epiphenomenalism is not to hold, it must be that Sam does what he does—responds ‘yes’ to the question ‘do you believe that all men are mortal?’—because he believes that all men are mortal, and because that belief has the content it does. As we have seen, however, there isn’t any *property* of believing that all men are mortal; there is only the *concept* believing that all men are mortal. But then according to Kim reductionism it can’t be that *believing that all men are mortal* plays a role in the causation of behavior: for according to Kim reductionism, it is by virtue of *having properties*, not by virtue of *falling under concepts*, that things or events are causally active. Of course this result is not confined to the property of believing that all men are mortal; it will hold equally for any belief. So we get the consequence *holding a belief with content C* is not causally efficacious; and that just is semantic epiphenomenalism.

Alternatively, suppose we think in terms of beliefs and their properties. Now according to Kim there isn’t any property of believing this or that, and hence also no property of having such and such a belief. But presumably there *is* the property of having such and such an H-belief. This H-belief will presumably be a neural event of some kind; and if it is the H-belief that all men are mortal, it must be characterized by *having as its content the proposition that all men are mortal*. Furthermore, it must be by virtue of having that content that this H-belief is causally active, if semantic epiphenomenalism is to be avoided. But *is* there any such property as *having as its (belief) content the proposition that all men are mortal*? According to Kim, presumably not. *Having as its (belief) content the proposition that all men are mortal* is presumably multiply-realizable, and hence functional; but according to Kim, there aren’t any functional properties. So it’s a concept. If semantic epiphenomenalism is to be avoided, then, it must be the case that Sam’s belief that all men are mortal causes what it does, makes its characteristic contribution to behavior, by virtue of its falling under the concept *having the content that all men are mortal*.

But Kim reductionism precludes this possibility; for it is only by having properties, not by falling under concepts, that a belief is causally active. But then it can’t be that the H-belief that all men are mortal makes its causal contribution to behavior by virtue of its falling under the concept *has as its content the proposition that all men are mortal*. And then it can’t be that

Sam does what he does, behaves as he does, because he has a belief with the content *all men are mortal*. And hence semantic epiphenomenalism is not avoided by Kim reductionism; it is instead an immediate consequence of it.

By way of reply someone might point out that (a) Sam has the physical property P_1 , the property of H-believing that all men are mortal, and that is by virtue of having P_1 that Sam does what he does, and (b) P_1 together with the presumably physical property *being human* entails *having a belief with the content that all men are mortal*; hence it is by virtue of having the physical property P_1 and *being human* that Sam does what he does; and *that* property entails *believing that all men are mortal*. Fair enough; but it doesn't follow that Sam does what he does by virtue of falling under the concept *has a belief with the content that all men are mortal*. It isn't the case that if P is causally active and P entails a concept or property P*, then P* is also causally active. Any property entails disjunctions; the latter, says Kim, are not causally active. Any property entails *has a property*; the latter, presumably, is not causally active. Semantic epiphenomenalism stands.¹³

C. Dualism of Properties/Concepts?

We may see the same thing in terms of a certain dualism of properties—or perhaps concepts: but for the moment let's ignore the distinction between properties and concepts. A property dualist, I take it, is someone who thinks there are both mental and physical properties, and that not all mental properties are identical with physical properties. (Some might add that the mental properties must be causally active). Now turn again to P_1 , the property of H-believing that all men are mortal. The conjunction of this property with the property of being human, we may suppose, entails other properties or concepts. In particular, it entails *believing that all men are mortal*, and it also entails N, the property of having a certain neural structure, the neural structure something must exhibit to have P_1 . P_1 therefore has at least a couple of other properties: being such that together with *being human* it entails *believing that all men are mortal* (call that property 'P*'), and that of entailing N. It seems reasonable to think of P*, the first property, as a mental property and of the second as a physical property. So P_1 has the mental property P* and also the physical property *entails N*. Now just as an event or substance may cause something by virtue of one of its properties,

¹³ I said that according to Kim, it is only properties that are causally active, not concepts. But suppose concepts *could* be causally active. Our question then would be whether the concept *holds a belief with the content all men are mortal* could be causally active. This concept, of course, does not correspond to any physical property. But then the very arguments Kim uses against nonreductive materialism, the causal exclusion argument, for example, would apply to this concept. So if Kim is right in his argument against nonreductive materialism, it can't be the case that the concept *believes that all men are mortal* is causally active.

but not by virtue of another, so a property can be causally active—such that by having it, an agent can cause something—by virtue of one of its properties and not by virtue of another. Thus the property *having an electrical charge* is causally active in certain ways, but not by virtue of its having the property of being a property, or being Sam’s favorite property. On Kim’s view it is by virtue of P_1 ’s having the property of entailing N that P_1 is causally active, not by virtue of its having P^* .

It is easy to see, I take it, that P_1 is causally active by virtue of its entailing N . It is because P_1 is such that whatever has it, displays structure N , that P_1 is such that whatever has it can make a causal contribution to behavior. But it is equally clear, I think, that on Kim’s way of looking at the matter, P_1 is not causally active by virtue of P^* . For if it were, then it would be by virtue of that entailment that P_1 causes behavior. That is, it would be because whatever has it and is human believes that all men are mortal, that whatever has it is causally active. But on Kim’s view, nothing is causally active by virtue of being human and believing that all men are mortal. So P_1 is not causally active by virtue of having P^* . What we have here, therefore, is a property (or perhaps concept) dualism—a second level property or concept dualism—that is reminiscent of the original problem: that apparently it is only by virtue of physical properties, and not by virtue of mental properties, that a belief is involved in the causal chain that leads to behavior.

D. Problems with Meanings

Now I’ve been assuming, in the above argument, that on Kim’s view there isn’t any such *property* as *believing that all men are mortal*; there is such a *concept*, all right, but no such property. But perhaps that’s importantly misleading. Let’s take a slightly different tack. Epiphenomenalism is false only if Sam behaves as he does because he has a belief with the content *all men are mortal*. Now consider the proposition,

- (7) Sam has a belief with the content *all men are mortal*.

I argued above that on Kim reductionism, (7) does not predicate a property of Sam, but only predicates of him that he falls under a certain concept. But perhaps that is mistaken. For there is of course the property *H-believes that all men are mortal*; this is a physical property which realizes *believing that all men are mortal* in human beings. Well, perhaps, in (7), what gets predicated of or ascribed to Sam is really the property *H-believes that all men are mortal*. If so, then presumably it could be that Sam behaves as he does because he believes that all men are mortal. More generally, when referring to a creature S of structure S' and saying “ S believes that all men are mortal” we are expressing a proposition that predicates *S'-belief that all men are mortal* of S . Perhaps this is how we are to understand Kim’s claim that, “In this way

multiply-realized properties are sundered into their diverse realizers in different species and structures, and in different possible worlds” (111).

Accordingly, perhaps the reduction is to go as follows: with respect to S_1 , the mental designator ‘believing that all men are mortal’ denotes physical property P_1 ; with respect to S_2 it denotes P_2 , and so on. If so, the locution ‘believing that all men are mortal’ displays a certain complex semantical character; it denotes different items on different occasions of its use. On some occasions of its use, ‘believing that all men are mortal’ denotes P_1 ; on others P_2 , and so on. When I am speaking of a member of S_1 , for example, then my sentence ‘Believing that all men are mortal is (identical with) P_1 ’ expresses a true identity proposition: on that occasion the mental property designator ‘believing that all men are mortal’ denotes the physical property P_1 . When I am instead speaking of a member of S_2 , that sentence expresses a falsehood, but ‘Believing that all men are mortal is (identical with) P_2 ’ expresses a truth—because on *that* occasion the locution ‘believing that all men are mortal’ denotes P_2 . On this account, the mental property designator ‘believing that all men are mortal’ would indeed denote a property, and in fact it would denote many *different* properties. And then it would be true that *believing that all men are mortal* (speaking of Sam) causes behavior—because, of course, it is true that P_1 causes behavior, and the proposition expressed predicates of P_1 the property of causing behavior, a property it has. Could this be what Kim has in mind?

But is it a coherent suggestion? I doubt it. First, the suggestion isn’t worked out fully enough: there are the usual problems with general claims such as ‘Some who believe that all men are mortal fail to display a sufficient grasp of the brevity of human life.’ Second, it does seem wrong to propose that there is no relevant property shared by creatures of different structures who believe that all men are mortal. Suppose I suffer a stroke at t ; the part of my brain in which P_1 is exemplified is damaged and I no longer exemplify P_1 ; but some other part of my brain leaps into the breach and I continue to believe that all men are mortal, now displaying a physical property P_2 rather than P_1 . Is it really sensible to say that in speaking of me after t , predicating of me that I believe that all men are mortal, you are now ascribing to me a different property from the one you ascribed to me before t ? Is it sensible to say that I now no longer have the property you predicated of me when you correctly asserted, before t , that I believed that all men are mortal? I think not.

But there are further problems. The idea, I take it, is that we human beings grasp such propositions as (7), understanding them to at least some extent; we can assert this proposition, can see at least some of its implications, and the like. If these things were not so, we wouldn’t so much as be able to entertain or affirm or state the claim we are currently examining.

And if we can state this claim, it will have to be by way of using some language, in my case English. So the idea will have to be that the English sentence

(8) Sam has a belief with the content *all men are mortal*

expresses the proposition that Sam H-believes that all men are mortal. But how can this happen? The idea is that 'believing that all men are mortal' denotes different properties on different occasions of its use. But how does it manage a thing like that? By virtue of its *meaning*, obviously. Well, what is its meaning? One possibility, abstractly speaking, is that this phrase is something like a multiply-ambiguous proper name or rigid designator. Perhaps that term, like other proper names, either fails to express a property at all (as on some theories of proper names) or expresses an essence of whatever it denotes (as on other theories of proper names). Perhaps it gets associated with what it denotes by virtue of some kind of baptism. Could it be that 'holding a belief with the content that all men are mortal' gets its meaning in this way, and is therefore something like a multiply-ambiguous proper name?

I don't think so. First, there are difficulties about how this term could have acquired the status of a proper name: how could the relevant kinds of baptisms have been performed? But there is a more pressing problem here. If 'holding a belief with the content that all men are mortal' is a proper name of a physical property, then either that name doesn't express any property at all, or else (if it does express a property) it expresses an essence of the property in question (i.e., expresses essences of such properties as P_1 , P_2 , and their comrades). But then sentences like 'Sam believes that all men are mortal' will fail to express a proposition entailing that Sam believes that all men are mortal. That is because 'believing that all men are mortal', on a given occasion of its use, either expresses nothing at all or else expresses at most an essence of P_1 . But according to Kim, it is not the case that having that property entails believing that all men are mortal: something of a different structure could have that property and *fail* to believe that all men are mortal. It therefore follows that a thing can have P_1 without believing that all men are mortal; and of course the same thing holds for P_2 and all the rest. But clearly that sentence *does* express a proposition entailing belief that all men are mortal. So this suggestion doesn't take us anywhere.

B, therefore, is not a proper name of P_1 . Is there some other way in which it could manage to denote different properties in different species? What kind of singular term could do a thing like that? Well, such terms as 'the property we're thinking of', 'the property currently under consideration,' and 'that property' all denote different properties with respect to different occasions of use; and they are all implicitly or explicitly indexical. So perhaps our term is

somehow indexical. How? Presumably by virtue of the *functionalization* of mental properties—or, since, according to Kim there aren't any functional properties, by virtue of the functionalization of mental concepts. Of course it's not clear, here, how we are to think of concepts, but suppose we set that concern aside for the moment. So what kind of concept could it be that would be expressed by 'believing that all men are mortal', and which would be such that this locution denotes different properties in different species? And what would this locution mean?

The obvious initial answer is something like

- (9) the physical property that realizes *believing that all men are mortal* in *that structure* (indicating a given structure).

It is hard to see how the relevant concept could fail to be equivalent to this one. But there is a two-fold problem. First and most important, we are trying to give or explain the meaning of the expression 'believing that all men are mortal'; but of course that expression occurs in (9). There is therefore a kind of vicious circularity in this effort to explain the meaning of that term. It's like saying "Here's how to understand the term 'Sam': it denotes whatever has Sam's essence." And second, even apart from the circularity, it couldn't really be that (9) gives us the meaning of B. For suppose it were the case that B just means (9). Then the sentence

- (10) If Sam believes that all men are mortal, then Sam has the physical property that realizes *believing that all men are mortal* in *that structure* (gesturing towards Sam)

would be true, express a true proposition, and would do so just by virtue of its meaning. It would be like "All bachelors are unmarried": one couldn't both understand it and also deny it. But clearly one can. For, one might not think there *is* any physical property that realizes *believing that all men are mortal* in that or any other structure. I myself don't believe that there is such a property, and I venture to say that I understand (9) about as well as the next person.

Of course there are other possibilities, here, for a way in which this term could behave in the fashion suggested; but none I can think of is at all plausible.

E. Kim Reductionism and the Probability of R

It is now time to return to our initial concern with the evolutionary argument against naturalism and P(R/N&E). I began by pointing out that

P(R/N&E&SE (semantic epiphenomenalism)) is low

and claimed that $P(\text{SE}/\text{N}\&\text{E})$ (N construed as including physicalism or materialism) is high. I then mentioned a couple of ways in which one who accepts N&E might try to avoid SE and considered one of them, Kim Reductionism, in more detail. So far I've been arguing that there are overwhelming problems with Kim Reductionism; in fact one can't (or anyway I can't) find a coherent program of reduction here. But now in conclusion I want to argue that even on the assumption that the Kim program works, $P(\text{R}/\text{N}\&\text{E}\&\text{Kim Reductionism})$ is low. The argument goes as follows. According to Kim Reduction, there is no such thing as the property *believing that all men are mortal*, but there is the *concept* of believing that all men are mortal. This concept is equivalent, in some unspecified sense, to a certain functional concept linking sensory input with behavioral output. And most crucially, for each species or structure S, there is the physical property *S-believing that all men are mortal*: call it 'P'. P is a physical, perhaps neurophysiological, property; but it also has propositional content in the sense that whatever has it, believes a certain proposition. Perhaps the connection between P and that content is a matter of causal or natural necessity, or perhaps it's a matter of broadly logical necessity; it doesn't matter for present purposes.

Now presumably the explanation of there being organisms that have this property P, with its particular content, is that having that property, with its specific neurological features and its role in causing behavior, has proved adaptive for those organisms. Displaying those features necessarily brings with it a certain propositional content, in this case the belief that all men are mortal. But now consider any particular physical property P with propositional content C. What we know about P is that (assuming it's not a spandrel) it has proved of survival value to have P. But that gives us not the faintest reason to think that the propositional content of P is true, or true at a time, or anything of the sort. Natural selection selects for those neurological properties; it must just take pot-luck when it comes to the content associated with them. P has the propositional content it has: there is no explaining why *that* property has that propositional content; it is just inexplicably attached to it. And the main point here is just that the fact that P is adaptive doesn't make it probable that the content associated with it, the proposition that all men are mortal, is true. Indeed, the best estimate of the probability that a given such content should be true, one thinks, can't be far from 1/2. That means it is very unlikely, with respect to a given person, that the proportion of her true beliefs to her false beliefs should be anywhere nearly as high as is required by reliability. Hence $P(\text{R}/\text{N}\&\text{E}\&\text{Kim Reductionism})$ is low. And hence even if, contrary to what I have argued, it successfully dispels the specter of epiphenomenalism, Kim Reductionism doesn't help resist the evolutionary argument against naturalism.