

Why Jaegwon Kim's Physicalism is Not Near Enough

An Implicit Argument for a non-Cartesian Interactionism

Part I

David Scharf

Abstract

Kim's argument for a minimal physicalism breaks down in two critical places. First, his functional reduction of cognitive/intentional properties fails to do justice to what we mean when we refer to belief, desire and the like, because it fails to account for their referential and normative content. And although Kim believes that he is saving what is distinctive about our cognitive/intentional life, by providing for mental causation in the context of functional reduction, it can be shown that mental causation without autonomy is inadequate, and irredeemably so, as Davidson anticipated. Thus Kim saves mental causation in name only. Although Kim has famously and, I believe, successfully critiqued Davidson's solution to the problem of mental autonomy in terms of nonreductive physicalism, Kim has never adequately addressed Davidson's motivation. Second, his minimal physicalism depends on a metaphysically fundamental partition between phenomenal consciousness and cognitive/intentional properties. This idea cannot be made coherent, because consciousness itself is intrinsically cognitive. This second argument will be developed in Part II. In introducing Part I, we will explain a crucial ambiguity in Kim's principle of the causal closure of the physical. By failing to appreciate the importance of this ambiguity, Kim's strong version of causal closure, which is anything but innocuous, is misleadingly made to appear so. In effect, Kim's under-appreciation of the complexity of the closure principle leads him to attack a straw man—in the form of an outmoded conception of immaterial minds—and ignore the possibility of an interactionism involving quantum neuroscience.

Key Words: Jaegwon Kim, physicalism, supervenience, interactionism, philosophy of mind, functional reduction, eliminative materialism, Donald Davidson, William James

NeuroQuantology 2010; 3: 337-353

Corresponding author: David Scharf, Ph.D.

Address: Maharishi University of Management, Department of Physics. Fairfield, Iowa. USA.

Phone: + 978-424-1828

e-mail: dscharf108@gmail.com

Submitted for Publication: September 6, 2010;

final revision received September 16, 2010; accepted September 21, 2010.

Introduction

In Jaegwon Kim's world view, "the basic parameters and constraints for the debates" about the nature of the mind take for granted "a set of broadly physicalist assumptions and aspirations that still guide and constrain our thinking today." And although none of the theories proposed based on these assumptions has proven satisfactory, "almost all the participants in the debate have stayed with physicalism, and ... have continued their allegiance to a physicalist worldview. Through the 1970s and 1980s and down to this day, the mind-body problem—*our* mind-body problem—has been that of finding a place for the mind in a world that is fundamentally physical." (Kim, 1998, p.2).

Not being a physicalist, this is not *my* mind-body problem and, in particular, the causal closure of the physical domain (I'll discuss that at some length below) is not as compelling to me (in the way it is usually understood) as it is to some of the other participants in the debates. Indeed, much of my motivation is to lay the groundwork for an interactionist approach to the mind-body problem and the need to develop a modern, scientifically refined conceptual model of what such an approach would look like. Although there are compelling reasons to reject a Cartesian interactionism of radically distinct substances, a more sophisticated interactionism is not only defensible, but is suggested by the present critical analysis.

More basic constraints on our theory of the mind than a commitment to physicalism, to my way of thinking, are that our theory should give sufficient recognition to the capacity for rationality and coherence of the mental domain and should acknowledge the fundamental, irreducible standing of consciousness. Kim's approach—which he characterizes as *minimal physicalism*—is detailed in his recent books, *Physicalism, or Something Near Enough* (2005) and in a more pedagogically tolerant presentation, *The Philosophy of Mind* (2006). His minimal physicalism incorporates four principal theses:

1. The causal closure of the physical.
2. The supervenience of the mental on the physical.

3. A fundamental partition between qualia and intentional/cognitive properties.
4. The proposal that qualia themselves are epiphenomenal.

In this discussion I intend to argue that all four theses are mistaken. But, though I believe that Kim's defense of physicalism misfires, his ruling out of nonreductive physicalism strikes me as correct, as I will soon explain. Although he presents his work as a defense of physicalism, what he has really shown is a disjunction—in order to account for mental causation, we must either accept a fully reductionist physicalism *or* allow that the mental domain is metaphysically fundamental and independent. Kim, obviously, chooses the former alternative but, as I hope to show, the evidence weighs heavily against this stance.

So how does Kim address our strong intuitions about the autonomy and independence of the mental domain? He accommodates these intuitions by radically distinguishing *consciousness and qualia* from *cognitive properties*, allowing pure subjectivity a place, but not in the context of the physical. These considerations box Kim in, forcing him to choose between mental causation and mental autonomy and leading him to bifurcate the mental domain into two parts which have almost nothing to do with each other—one part which is causally efficacious and one part which is ontologically autonomous (though causally epiphenomenal).

Kim's argument breaks down in two critical places. First, his functional reduction of cognitive/intentional properties fails to do justice to what we mean when we refer to belief, desire, representation and the like. Donald Davidson is justly celebrated for staunchly defending the autonomy implicitly embedded in these concepts. And, although Kim believes he is saving what is distinctive about our cognitive/intentional life by providing for mental causation, in the context of functional reduction, I will try to show that mental causation without autonomy is inadequate, and irredeemably so. Second, Kim's minimal physicalism depends on not just a pragmatic distinction between consciousness (or qualia) and cognitive/intentional properties, but a

metaphysically fundamental partition between the two. This idea cannot be made coherent, for consciousness itself is intrinsically cognitive. An extensive body of contemporary research in cognitive neuroscience has documented the principle, anticipated by William James that, “The particulars of the distribution of consciousness, so far as we know them, point to its being efficacious.” (James, 1890, p. 138).

I expand these ideas in the two parts of this article²:

I. Explains the autonomy and irreducibility of normative cognitive/intentional properties.

A. The origins of minimal physicalism—outlines the development and motivations for Kim's position, with particular reference to his well-argued and convincing refutation of nonreductive physicalism.

B. Makes the case that a robust autonomy is implicit in our conception of the cognitive/intentional domain.

II. Examines the relationship between phenomenal consciousness and cognitive/intentional properties, and critiques Kim's tenet that qualia are devoid of cognitive/intentional import.

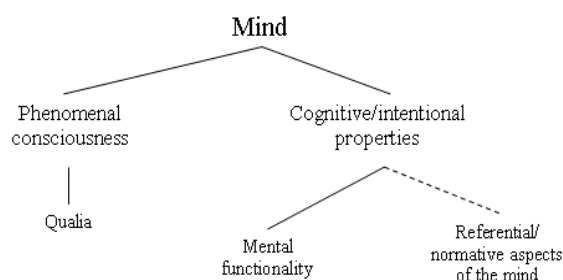


Figure 1. Analysis of mind implicit in Kim's minimal physicalism.

By way of introduction, and with reference to Figure 1, here is a brief sketch of Kim's minimal physicalism and the main threads of my critique. The first, and most

important, tenet at the basis of Kim's analysis of mind is the sharp distinction between phenomenal consciousness and what he describes as “cognitive/intentional” properties. By phenomenal consciousness he means primarily—and perhaps even exclusively—sensory qualities, i.e. qualia. By cognitive/intentional properties he encompasses beliefs, desires, hopes, wishes and other psychological states that can be expressed in terms of propositional attitudes. There are two themes which motivate his sharp distinction between qualia and cognitive/intentional properties. First is the *explanatory gap*—the so-called “hard problem” of consciousness—which points to the irreducibility of phenomenal consciousness on the grounds that, even after neuroscience has completely mapped the brain and documented its biophysical processes, we still will not have explained subjective experience. Second is mental causation. Kim proposes that we can acknowledge the explanatory gap *and* provide for mental causation by sharply differentiating qualia from cognitive/intentional properties: Qualia are irreducible but causally irrelevant (his nod to the explanatory gap), while cognitive/intentional properties are to be understood in terms of their function, in other words, in terms of their causal role in mediating between sensory input and behavioral output. Functional reduction, then, will provide the basis for mental causation.

There are two main threads to my critique. First I don't believe that Kim has adequately addressed the referential/normative aspects of the mind. In Figure 1, a dashed line is used to indicate the link between cognitive/intentional properties and their referential/normative characteristics because these characteristics belong there, but they're largely missing from Kim's analysis. Davidson has famously argued that the rationality and coherence of the mental domain implies its autonomy and irreducibility, and the first thread of my critique is intended to follow up on his insights.

Second (I develop this thread in Part II) I argue that Kim's radical differentiation between phenomenal consciousness and the cognitive/intentional domain is unworkable

² Part II is forthcoming in NeuroQuantology, December 2010; 8(4).

and, in fact, incoherent. By undermining Kim's radical demarcation, I am indirectly providing support for a celebrated argument for interactionism, put forward by William James more than a century ago. Paraphrased and abbreviated, James's argument is that, if consciousness is both irreducible and functional, then it must (as a nonphysical sort of thing) interact with the brain. Contemporary cognitive neuroscience has provided compelling evidence that consciousness is functional and, indeed, adaptive. Hence, if the consciousness that the explanatory gap says is irreducible—phenomenal consciousness—is the same as the consciousness which cognitive neuroscience tells us is functional—cognitive/intentional—then William James is right, and interactionism is a reasonable point of view. Both threads of my critique funnel toward the conclusion that interactionism needs to be taken seriously, where consciousness, and at least some cognitive/intentional features, are metaphysically fundamental and interact with the brain.

In Part I, we will look at Kim's refutation—which I consider to be successful—of nonreductive physicalism. Although there are several variations of nonreductive physicalism, including the early twentieth century theory of emergence, dual-access theories and the anomalous monism of Donald Davidson, Kim has shown that they all involve a similar shortcoming: They either violate the causal closure of the physical or they fail to provide for mental causation. In other words, they either are not *really* physicalism, after all, or they render the mental domain causally inert, epiphenomenal. Anomalous monism was Davidson's attempt at a solution for the issue of the rationality and coherence of our mental life. The main point which I hope to bring out from Part I is that, although Kim has undermined Davidson's *solution*, he has not satisfactorily addressed Davidson's *motivation*.

The problem, as Kim sees it, has to do with reconciling the causal closure of the physical with the principle of causal exclusion. Kim formulates these principles as follows (Kim, 2005, p.15-17):

The causal closure of the physical domain. If a physical event has a cause at t , then it has a physical cause at t .

Principle of causal exclusion. If an event e has a sufficient cause c at t , no event at t distinct from c can be a cause of e (unless this is a genuine case of causal overdetermination).³

In other words, to Kim, physical events have physical causes and this excludes their having non-physical causes. From this it would follow that mental events must be physical if they are causally efficacious in the physical world. The flaw in this argument has to do with causal closure, whose seeming innocuousness rests on an ambiguity in the term *physical*₁:

*physical*₁ = belonging to the world of sense perception, or analogous to the objects or events of sense perception.

*physical*₂ = belonging to the domain of physics.

*Physical*₁ is the pre-scientific conception, and is at the basis of most of our semantic intuitions about the physical or material world. But causal closure is not obvious—indeed it's quite problematic—if by *physical* we mean *physical*₁; and something like this is just what most physicalists, including Kim, do mean. Accordingly, it would be hard to defend the closure principle if it is interpreted as:

(Type A closure) If a physical event has a cause at t , then it has a cause at t that either belongs to the world of sense perception or is analogous to objects or events of sense perception.

³ Kim makes a convincing case that mental/physical causation is not an instance of over determination and we will not pursue this line of argument.

⁴ Herbert Feigl (1958) introduced a similar distinction where he construed *physical*₁ in terms of "the type of concepts and laws which suffice in principle for the explanation and prediction of inorganic processes."

And yet causal closure is practically tautologous if by *physical* we mean *physical₂*. It is practically tautologous that

(Type B closure) If a physical event has a cause at *t*, then it has a cause at *t* belonging to the domain of physics.⁵

Why is this obvious? Consider that a hundred years ago, the only two forces that physics acknowledged were electromagnetism and gravity. Then, as a result of nuclear investigations, two new forces were incorporated into the canon of fundamental physics—the weak and strong nuclear forces. The sociological moral is that physicists as a group are a pragmatic sort and whenever they feel the need to introduce new particles, forces or fields into the fundamental canon, they will do so. So, if it turns out that awareness, cognition, intentions and so forth causally affect the firing of the neurons in a way that cannot be explained on the basis of today's physics, physicists will not hesitate to account for mental causation by introducing new forces, fields and so on as needed. The domain of physics is indefinitely expandable.

Kim, and most physicalists, unwittingly rely on the plausibility of Type B closure to gain acceptance for the much stronger assumptions inherent in Type A closure. But once implicit acceptance has been gained for Type A closure, the assumptions inherent in Kim's physicalism (such as "bits of matter" and supervenience, to be described below) acquire more seeming plausibility than they deserve.

Of course, one might naively believe that *physical₂* is simply the scientific extension of *physical₁*. So, for example, although theoretical entities like atoms and sub-atomic particles are too small to see, we would still consider them to be physical by analogy to sensory objects such as billiard balls and marbles. But the domain of physics is not a straightforward extension of our familiar sensory-based world; and its laws

differ greatly from the familiar principles of classical physics.⁶

For example, a number of distinguished physicists have argued that the role of the observer in relativity theory and in quantum measurement involves essential and irreducible reference to consciousness. If they are right—and for the sake of this argument, let's say they are—consciousness belongs to the fundamental canon of physics and is not reducible to brain chemistry. Moreover, if—to consider quantum measurement—consciousness is what causes the collapse of the wave function to an eigenstate of an observable, then quantum measurement is incompatible with physicalism. While this view of quantum measurement is controversial, the important point to note here is that merely including consciousness or mental phenomena generally, in the domain of physics would not thereby satisfy Type A causal closure and would not be sufficient for physicalism.

So what view of physics does the physicalist require? Before we attempt to answer in the abstract, let's look at the view of physics that Kim endorses. Many physicalists would agree with him that the world is composed of what he refers to as "bits of matter," causally interacting against the stable background of space and time. This is essentially the viewpoint of classical physics and, from Kim's perspective, nothing has significantly changed insofar as the debates on the mind-body problem are concerned. Here's how he puts it:

On the overall shape and makeup of the world in essential outlines, we must depend on what physics, our fundamental science, tells us. I believe that the broad basic features of the world as described by modern physics, what is intelligible and is

⁵ Based on Type B closure, one could hold that mind-brain interactionism can (and should) be modeled in terms of physics, without being committed to the idea that mental phenomena are physical or material in the traditional senses of these terms (*physical₁*).

⁶ In associating classical physics with the principles and categories implicit in the familiar world of sensory-based experience, I am essentially linking *physical₁* to classical physics. This evokes Niels Bohr's emphasis on the extent to which quantum physics has distanced itself from the familiar pictorial representations so indispensable to classical physics. Strictly speaking, though, classical physics had already dissociated itself from some of the assumptions that seem embedded in an uncritical, sensory-based view of nature. The abandonment of the geocentric worldview, and the unification of celestial and terrestrial mechanics by Galileo and Newton, are early instances of this. Nevertheless, compared to quantum physics and, especially, unified quantum field theories, it is reasonable to associate classical physics—with its pictorial models and analogies, and familiar, sensory-based assumptions about space, time, matter and causation—with *physical₁*.

of interest to those of us who are not science specialists, has been relatively stable through the flux of changing physical theories, and this is what forms the background of the debates on the mind-body problem. (Kim, 2005, p.149)

Of course Kim realizes that no physicist would agree that “the broad basic features of the world as described by modern physics” have been “relatively stable through the flux of changing physical theories” Physicists have long prided themselves on the worldview-shattering results of, first, relativity theory, and then quantum physics and, most recently, the integration of the two in unified quantum field theory. Indeed these have been conceptual revolutions of staggering proportions. What Kim means to suggest is that none of these revolutions will affect our understanding of mind and its relationship with matter. That means the functionalist reduction Kim endorses requires only the billiard-ball physics conception of classical physics. In the following passage, taken from *Physicalism, or Something Near Enough*, he makes it clear that the view of matter embodied by classical physics is all that he regards as relevant to the reduction of mind:

The core of contemporary physicalism is the idea that all things that exist in this world are bits of matter and structures aggregated out of bits of matter, all behaving in accordance with laws of physics, and that any phenomenon of the world can be physically explained if it can be explained at all. (Kim, 2005, pp.149-150)

And in his textbook, *Philosophy of Mind*, Kim again emphasizes his view of the physical reductive base as consisting of “bits of matter.”

As we will see, there are various forms of both property dualism and reductionist physicalism. However, they all share one thing in common: the rejection of immaterial minds. Contemporary property dualism and reductive physicalism acknowledge only objects of one kind in the world—bits of matter and increasingly complex structures aggregated out of bits of matter. (Kim, 2006, pp.51-52)

My point, in underlining Kim's reliance on the conception of matter embodied in classical physics is this: Implicit in the classical “bits of matter” worldview

(characterized as *physical_i* above) are several assumptions which Kim is relying on to undergird physicalism. Bits of matter are conceived as consisting of spatially localized, independently existing particles which interact causally against the backdrop of space and time. What's more, their causal interactions depend on spatiotemporal contiguity—local contact—and wholes are understood microreductively, where the whole is nothing but an arrangement of the parts. In Kim's presentation these assumptions are encapsulated in his concept of *supervenience*.

Kim makes a point of clarifying the type of mind-body dependency relationship that he has in mind in terms of supervenience, where supervenience requires type-to-type correlations between mental events and neurophysical events, such as groups of neurons firing. Kim's supervenience of the mental on the physical involves an asymmetric dependence. As he depicts it, “... what happens in our mental life is wholly dependent on and determined by, what happens with our bodily processes” (Kim, 2005, p.14).

Kim's physicalism, then, embodies deep and, to my mind, deeply problematic assumptions concerning the *level* of nature's functioning at which mental phenomena exist. Do they belong to the fundamental physical canon, or are they expressed only at a more surface level of existence, say at the level of molecular biology? In other words, are intentions, reasons and the like fundamental and irreducible? Or are they dependent and derivative—supervenient—as Kim suggests? Denying metaphysical primacy to the mental is the real motivation for physicalism, as well as for its central thesis, the (Type A) causal closure of the physical.

While it is common knowledge that physics has discarded the “bits of matter” worldview implicit in the classical perspective, the character of advanced physics is not readily accessible to the non-specialist, and there is no sufficient consensus about its interpretation to easily draw definite conclusions for a philosophical issue such as the mind-body problem. Hence it is not surprising that the default position of most philosophers of mind has been to

simply dismiss the relevance of advanced physics and to rely on an essentially classical worldview.

However we should emphasize that most of the assumptions undergirding supervenience are no longer held to be true at the level of fundamental physics. Contemporary thinkers tend to be familiar with the features of wholeness and nonlocality inherent in spatial (EPR) entanglement, where the states of spatially separated particles cannot be independently specified. A single wave function describes the state of the whole and a measurement affects the state of not only the particle directly measured, but the state of the other particle(s) as well. Fewer of them are familiar with temporal entanglement, in which a measurement on a particle *now* can affect its state *in the past*. In previous work (Scharf, 1989) I explored the microreductive unanalyzability of quantum measurement, and similar considerations would show that quantum measurement is not supervenient on the microstructure of the measurement apparatus. Moreover, advances in contemporary quantum field theory involve stages of unification that are successfully eliminating categories and distinctions that are deeply embedded in our sensory based, "bits of matter" conception of physics.

So it strikes me that Kim is overconfident in dismissing the significance of these considerations from advanced physics for the mind-body problem. For example, if we weren't bound by classical physics assumptions about causation and contiguity, would we be as willing to dismiss our common-sense, direct realist intuitions about perception in favor of the brain-in-a-vat sort of scenario that has so dominated and perplexed contemporary philosophy of mind? If all the theoretical resources of contemporary physics were available, a straightforward account of perception might plausibly suggest a nonlocal entanglement between the brain and the environment.⁷

⁷ As noted by John Austin and more recently Hilary Putnam, veridical perception may employ fundamentally different physical mechanisms than illusion and hallucination. What I'm suggesting is that quantum nonlocality may offer a way to start exploring what these mechanisms might be. While speculative, these considerations put the burden of proof on Kim and like-minded physicalists to justify excluding advanced physics from the philosophy of mind. Incidentally, while we're in a speculative

An unprejudiced specification of the structure and function of consciousness and cognition is likely to include features of wholeness, nonlocality and entanglement that are lost—eliminated—if we try to force a reduction of mind to brain. The fact that it has proven so difficult to find a satisfactory reductive theory of mind despite decades of intense effort may well be because the physical correlates of consciousness and cognition involve features of advanced physics not available in the classical worldview in which supervenience features prominently. It would take us too far beyond the present scope of this discussion to develop these themes; although in my concluding section (of Part II) I will indicate the direction my own thinking has taken. But by providing a systematic critique of Kim's minimal physicalism I hope to lay the groundwork for why we need to consider far-reaching new approaches.

A. The Origins of Minimal Physicalism: The critique of non-reductive physicalism, the rescue of mental causation, and the acknowledgement of the irreducibility of phenomenal consciousness

Obviously there are causal relations between mental and physical events. In sensation and perception our sensory qualia and cognitive beliefs are influenced by physical stimulation mediated by the nervous system. And in acting for reasons, our beliefs and desires affect our behavior, again as mediated by our nervous system. Kim certainly takes the existence of mental/physical causal interactions as a fundamental commitment and this, together with his commitment to physicalism, drives his reasoning.

Autonomy of the mental vs. physicalist mental causation

Donald Davidson defends what he characterizes as a kind of physicalist position—namely that in order to account for mental causation we must suppose that

mode, parallel considerations suggest that quantum *temporal* nonlocality might undergird a direct-realist approach to veridical memory and even precognition, should that prove to be a genuine phenomenon.

mental events are physical—but he combines his version of physicalism with an additional imperative. For Davidson, considerations of rationality and coherence dictate that mental events, as they fall under mental types, must be anomalous, meaning that they cannot be placed within the network of physical causal laws as instances of mental types. His anomalous monism incorporates token physicalism, according to which individual mental events are physical events, but mental types do not reduce to physical types. Davidson is trying to do justice to our intuitions regarding the autonomy of the mental domain, while at the same time providing for mental causation within the physicalist paradigm.

Davidson emphasizes the importance of an insight of Immanuel Kant's—indeed it is one of the chief organizing principles underlying Kant's philosophy. Here's a succinct statement of Kant's perspective on reason, illustrating the fundamental role of autonomy for our conception of rationality:

We cannot possibly conceive of a reason as being consciously directed from outside in regard to its judgments; for in that case the subject would attribute the determination of his power of judgment, not to his reason, but to an impulsion. Reason must look upon itself as the author of its own principles independently of alien influences. [*Groundwork of the Metaphysic of Morals* 101=448] (Kant, 1785/1964).

Kant is explaining that there is something inconceivable (paradoxical, incoherent) about a rational being (he means us) that is “consciously directed from outside”—in other words, that is being directed from outside and is conscious of being so directed—in regard to its judgments (cognitive/intentional life). Such a being would regard its judgments as being determined by impulse rather than reasons.

Both Kant and Davidson express the constraints of rationality and coherence in terms of autonomy—which is to be contrasted with being determined by something external.

Why cannot our cognitive mental life—as instantiating mental properties—be supervenient on the causal flow of physical events, governed by physical laws? Physical

laws have their own imperatives, which are under no constraint to support the requirement that the flow of our cognitive life should make sense. If we believe that we are (or can be) rational and coherent it follows that we should believe that we are (or can be) free to respond to the requirements of logic and evidence. This seems to capture a deep and fundamental intuition about who we are and the nature of mind: rationality implies autonomy, and autonomy implies freedom to respond to the normative principles of logic and evidence, objective moral standards, and so forth.

Davidson represents this freedom to respond in terms of *anomalousness*. He thinks that there are no strict psychophysical laws, or strict psychological laws of any kind. I don't think Davidson is right about this—at least some of our normative principles have already been formulated in terms which are quite strict and precise, and this precision may characterize descriptive (not simply prescriptive) laws of thought, under suitable conditions. But for now, let us turn to Kim's critique of Davidson, which is aimed at his anomalous monism, but does not address Davidson's core motivation to provide a realistic and workable account of autonomy.

Kim's critique of nonreductive physicalism

This freedom is what Davidson was trying to characterize with his doctrine of anomalous monism. Kim's criticism, however, shows that anomalous monism has the unanticipated consequence of rendering the mental domain causally inert and epiphenomenal. In Kim's words:

To be sure, anomalous monism is not epiphenomenalism in the classic sense, since individual mental events are allowed to be causes of other events. The point, though, is that it is an epiphenomenalism of *mental properties*—we may call it “mental property epiphenomenalism”—in that it renders mental properties and kinds causally irrelevant.... Mental properties play no role in making mental events either causes or effects. To make this vivid: If you were to redistribute mental properties over the events of this world any way you please—you might even remove them entirely from all events, making all of them purely physical—that would not alter, in the slightest way, the network of causal

relations of this world; it would not add or subtract a single causal relation anywhere in the world! (Kim, 2006, pp.188-9)

In his attempt to safeguard the autonomy of the mental domain, Davidson has done more than he intended, with the result that he has sidelined mental properties altogether. The lesson we can draw from Kim's critique is that we recognize mental events as causally efficacious *in virtue of their mental properties*, and our theories should reflect this. Hence token physicalism does not provide an adequate account of mental causation.

Kim is right, I believe, and in general his critique of nonreductive physicalism in all its variations is right for the same reason. In trying to preserve the independence of the mental domain in the context of a physicalist framework the nonreductive physicalist ends up sidelining mental properties. If it is in virtue of physical properties (location in space and time, mass, charge and the like) that mental events are causally efficacious, then their unreduced mental properties play no causal or explanatory role and are thereby rendered epiphenomenal—unsatisfactory for a meaningful philosophy of mind. We are left with this alternative: reductive physicalism or unreduced mental causation.

The dual-access theory of Herbert Feigl has been receiving quite a bit of renewed attention recently. In a classic paper, Feigl (1958) proposed that mental properties are identical to neurophysiological processes in the brain and that, rather than being a duality of properties, this represents a duality of epistemic access. That is, we may become aware of our mental life introspectively, by direct awareness or indirectly, through the external investigations of neuroscience. So our first-person reports and the third-person scientific descriptions of the functioning of our brain processes may represent two different ways of accessing or describing the same phenomena.

I won't attempt to rehearse the long tradition of commentary on Feigl's important idea. But, following Kim, I think that Feigl's dual access theory suffers from the same shortcoming that undermines Davidson's anomalous monism, namely that the neurophysiological processes—

representing the physical side of the dual access—are embedded in the broader context of the physical processes of the inorganic physical universe. These physical laws have their own imperatives regardless of the associated mental aspects. Hence it is difficult to avoid Kim's conclusion that nonreductive physicalism renders the mental epiphenomenal and causally irrelevant. The causal nexus that Feigl has in mind is that of classical physics where causal determination is transmitted locally and mechanistically, rendering any normative considerations—that the mental side of the equation might be sensitive to—causally irrelevant and thereby depriving our mental life of its essential cognitive autonomy.

To get a sense of how deeply problematic epiphenomenalism is for nonreductive physicalism, consider a recent development, due to Dempsey and Shani (2008), in which they revisit Feigl's dual access theory, "emphasizing the capacity of autonomous systems to control their own micro-constituents for the benefit of the whole." Like Feigl, they "embrace a thoroughgoing physicalism, one which (type) identifies conscious phenomena with certain neurophysiological phenomena." Dempsey and Shani, however, explicitly reject Kim's supervenience and its microreductive commitments. Their model conceives a sentient organism as a dynamical system, which is a self-organizing system in which the organism's requirement for maintaining stability involves downward causation where the micro-components of the system are constrained to function so as to maintain the stability of the whole.

In appreciating that an adequate account of the mental domain will necessitate our abandoning supervenience, Dempsey and Shani's model represents an advance, but with two qualifications. The first qualification has to do with the incompleteness of their theory, and is basically my suggestion for where further clarification is needed. But the second qualification is, I believe, a fatal criticism of their model, as it currently stands.

First, then, supervenience is essential to Kim's understanding of (Type A) causal closure and, insofar as downward causation violates supervenience, Kim would not

regard their model as authentic physicalism. But more crucial, from my perspective, is the question of whether Dempsey and Shani are consistent with even the much more tolerant Type B causal closure. In other words, is their conception of downward causation compatible with contemporary physics and, if not, then what sort of new physics will be required, and how will they integrate the principles of self-organizing systems into the broader nexus of physical theory. If self-organizing, biological systems constrain the behavior of their constituent particles through downward causation, physics is going to want to understand how this works.

My second qualification is that Dempsey and Shani have not gone far enough: by rejecting supervenience of biological, self-preserving systems they are safeguarding biological autonomy, but they still have not provided for cognitive autonomy. Their proposal still identifies consciousness with neuro-biological processes and, while their model allows these a primitive sort of autonomy, corresponding to biological self-preservation, one would have to believe that this is sufficient to account for the cognitive autonomy intrinsic to consciousness. Their model still leaves consciousness supervenient on biological processes, and the autonomy inherent in our capacity to respond to the normative principles of ethics and rationality cannot be adequately accounted for on the basis of biological self-preservation or, more generally, on the basis of the self-organizing features of dynamical systems.⁸

Thus, as long as the organism avoided predators and other dangers, acted to acquire nutrition and, in general, behaved as a self-preserving biological system, the description of its neural processes engendered by mental access would be irrelevant. To make the criticism vivid, as long as our neural processes were getting us

out of the way of the charging rhinoceros, the mental-access side of the equation might just as well have us believe we were pleasantly and safely at home with our family. Kim's incisive criticism of Davidson's non-reductive physicalism is equally applicable to Dempsey and Shani: "If you were to redistribute mental properties over the events of this world any way you please—you might even remove them entirely from all events, making all of them purely physical—that would not alter, in the slightest way, the network of causal relations of this world" Dual access theory grounded in physicalism, like property dualism or substance dualism based on physicalism, renders the mental domain epiphenomenal—even if it's physicalism at the biological level (Dempsey and Shani) rather than the level of microphysics (Davidson).

To summarize Kim's critique of non-reductive physicalism, as long as the mental/physical events are embedded in the larger, physical nexus, the mental side of the equation will be regarded as epiphenomenal. Kim's minimal physicalism addresses this by saying: OK, we'll allow that phenomenal consciousness *is* epiphenomenal, but all the cognitive aspects of mind reduce to the physical. But Kim's critique is more powerful than he realizes, because it renders the *normative aspects* of cognition to be epiphenomenal, as well, even if the cognitive processes themselves are physical—with intolerable and paradoxical consequences. So, as I will explain in greater detail in the following section, Kim's critique of non-reductive physicalism is equally devastating when applied to his own proposed solution, minimal physicalism.

From my perspective, cognitive autonomy is nonnegotiable, because it is essential for our concepts of rationality and moral responsibility. My preferred approach to reconciling mind with the domain of physics, is to consider that if we combine the requirement for a robust conception of cognitive autonomy with the requirement for Type B causal closure, we will end up incorporating autonomous cognitive principles into the canon of fundamental physics.⁹ It will be beyond the scope of the

⁸ Epiphenomenalism (of the mental access side of the dual access equation) would fatally undermine any proposal to account for the normative character of mind on the basis of biological self-preservation. Thus, any appeal to causal feedback mechanisms, such as natural selection or stimulus/response learning, will fail because the mental access side of the equation is incidental to the causally efficacious aspect, which is physical. This point is reminiscent of William James's famous argument for interactionism (James, 1890, p. 138).

⁹ It might help to lay this out more explicitly:

present paper to develop these ideas at any length, but in my concluding remarks I will outline a kind of dual access, identity theory—but at the level of fundamental physics, not at the level of neurophysiology. In the meantime, it will help to motivate the need for a radically new slant along these lines, if we can show why Kim's approach—minimal physicalism—fails to provide for cognitive autonomy.

Minimal Physicalism

A successful approach to the mind-body question must at least provide for mental causation. But how does Kim propose to do this, and what does his proposal do to cognitive autonomy? To provide for the causal efficacy of the mental without jeopardizing his commitment to the primacy of the physical, Kim and like-minded physicalists *have to* say that the mental is in some sense causally efficacious through reduction to its underlying neural correlates. If Kim's argument against nonreductive physicalism is convincing—and I think it is—it follows that, if we are to preserve mental causation in the context of physicalism, then "... we had better be prepared to take reductionism seriously, whether we like it or not." (Kim, 2005, p.22)

Because of his acknowledgement of the irreducibility of subjective experience, Kim adopts a limited form of reductionism, where subjective experience involves consciousness, or the having of qualia, or where there is "something it is like," to use the description made popular by Thomas Nagel. (Nagel, 1974) Unlike qualia, intentional/cognitive properties are functionalizable, in Kim's view. That is, they can be analyzed in terms of their functional roles and behavioral outcomes. This sets

them up for reduction—identification with the underlying neural processes which have those same functional roles and which produce the subsequent behavioral outcomes, the assumption being that those neural processes are microreductively explainable in terms of fundamental physics. This idea gives the intentional/cognitive properties causal relevance by identifying them with their underlying neural correlates, while qualia are left as an epiphenomenal residue. Hence Kim is committed to an absolute, ontological distinction between qualia, which are not reducible, and intentional/cognitive properties which can be reduced to their underlying neural substrate.

Although Kim believes that he is not an eliminativist—he doesn't think that mental properties, particularly cognitive/intentional properties, are dispensable or that he is dispensing with them—he is actually teetering pretty close to the brink. Consider that Darwinian materialists, in the 19th century, argued that, by providing causal explanations for the apparently purposive functioning of biological organs and organ systems, they had eliminated the need for a cognitive/intentional understanding of biological systems in terms of vital forces, entelechies, or Divine intentions. Interestingly, Kim is using the same sort of argument—proposing to explain cognitive/intentional properties, after suitably functionalizing them, in terms of underlying physical causes—but drawing the opposite conclusion! Instead of eliminating cognitive/intentional properties he believes his functional/causal reduction is conserving them, allowing us to save at least this portion of the mental domain from elimination or epiphenomenalism. As we continue our examination of Kim's functional reduction, I believe it will become increasingly apparent that his position commits him to a more extreme form of eliminativism than he—or anyone—should feel comfortable with.

William James's *Principles of Psychology* provides elegant expression to the interactionist alternative to reductionism, and it will be a good idea to remind ourselves that there is another choice, before we look more closely at the shortcomings of Kim's position. In the

(Cognitive Autonomy) We are sometimes capable of responding freely to the normative requirements of ethical behavior and rational thought. This implies independence from the constraints of the physical domain, which has its own imperatives regardless of the normative implications.

(Causal Interaction) Obviously there are causal relations between mental and physical events. If some of our mental events are autonomous, this autonomy may be expressed in physical behavior.

(Type B Causal Closure) If a physical event has a cause at *t*, then it has a cause at *t* belonging to the domain of physics.

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

(Conclusion) The canon of fundamental physics includes autonomous cognitive principles.

following passage, James is defending the common sense, interactionist point of view when he speaks of consciousness as exerting pressure or inhibiting neural processes in the brain in accordance with its *interests*:

Well, just such pressure and such inhibition are what consciousness *seems* to be exerting all the while. And the interests in whose favor it seems to exert them are *its* interests and its alone, interests which it *creates*, and which, but for it, would have no status in the realm of being whatever. We talk, it is true, when we are darwinizing, as if the mere *body* that owns the brain had interests; we speak about the utilities of its various organs and how they help or hinder the body's survival Considered merely physically, all that can be said of them [the functions of the body] is that *if* they occur in a certain way survival will as a matter of fact prove to be their incidental consequence. The organs themselves, and all the rest of the physical world, will, however, all the time be quite indifferent to this consequence, and would quite as cheerfully, the circumstances changed, compass the animal's destruction.... Every actually existing consciousness seems to itself at any rate to be a *fighter for ends*, of which many, but for its presence, would not be ends at all. Its powers of cognition are mainly subservient to these ends, discerning which facts further them and which do not (James, 1890, pp. 140-141).

For James, interests are intrinsic to consciousness, and this seems self-evident to him. In contrast, Kim believes we can understand our interests and, indeed, our whole system of cognitive/intentional properties as separate and distinct from consciousness. Kim's suggestion that this system of cognitive/intentional properties could be, in James's words, "considered merely physically" and might be nothing over and above causal processes would, for James, seem conceptually incoherent. Causal feedback systems, such as Darwinian natural selection, would not improve James's assessment, since the assignment of value seems to involve an entirely distinct conceptual category from the functional/causal processes of biology and, ultimately, physics.

Kim would argue—incorrectly, I believe—that James's use of the term *consciousness* is ambiguous between

phenomenal consciousness (what it is like to be having subjective experience) and a cognitive/intentional agency with the functional resources for achieving its ends. Phenomenal consciousness, as Kim describes it, is passive and epiphenomenal. And cognitive/intentional agency, as Kim would argue in direct contradiction to James, can be reduced to purely causal, physical processes. In Part II, I will argue that—consistent with James's view—phenomenal consciousness has intrinsic cognitive/intentional significance. Kim would deny this—has to deny this—maintaining that cognitive/intentional properties are separable from phenomenal consciousness and can be functionally reduced to strictly causal processes. Now, let's consider whether he can really carry out this functional reduction without eliminating the essential core of the cognitive/intentional domain—its autonomy. As we will see, causation without autonomy leads to cognitive/intentional epiphenomenalism; this applies to Kim's functional reduction as much as it does to the nonreductive physicalist positions he has so astutely criticized.

B. Autonomy and the Irreducibility of Cognitive/Intentional Properties

Kim is under no illusion that his suggestion that cognitive/intentional properties can be functionally reduced to the neuronal processes that mediate between sensory inputs and behavioral outputs is going to seem intuitively plausible. Cognitive/intentional properties seem to be fundamentally *unlike* the causal properties that molecular biology describes. This unlikeness is not just relevant to whether we can reduce qualia. In other words, we cannot finesse the "hard problem" by merely quarantining qualia, as Kim recommends that we do—we have to address the unlikeness of cognitive/intentional properties as well as qualia.

Our Cognitive/Intentional Life seems Responsive to Normative Principles

Let's sharpen this sense of unlikeness. Intentionality involves reference to or representation of external states of affairs or abstract facts—an intentional state is *about* something. This aboutness, or reference,

does not seem to be microreductively analyzable in terms of supervenience, at least not in any direct way. If we were to look to causal interactions, no matter how complex or sophisticated they may be, to provide the supervenience base for cognitive/intentional states, there would be no way to directly map relations which are inherently nonlocal or abstract onto contiguous causal mechanisms which are inherently local and concrete. Intentionality and cognition would have to be recast in terms of what Kim calls an "internal core" which can be identified with functional processes in the brain. The rest (the objects of intention and cognition) might be causally related to this internal core but are not essential to it. In addition—and significantly—cognition takes place against a normative background, which also seems *prima facie* irreducible to cause and function. Patterns of thought and consequent actions can be judged for logical consistency, rationality and so forth, as well as for moral decency. Reference also is normative, in that we may succeed in referring or fail to refer.

And it is not simply that the rationality, coherence and decency of our cognitive/intentional life can be judged, after the fact. Our thinking responds to these normative principles—under amenable circumstances—or so common sense would have us believe. Normative principles seem to directly influence our knowledge, beliefs and desires. Kim himself advises us that, "We will do better to focus on its [common sense psychology's] normative role in the evaluation of actions and the formulation of intentions and decisions." (Kim, 1993, p. 263, n46) But normative considerations count *against* the kind of reduction to cause and function that Kim advocates. Most of us are inclined to agree with William James that, even after Darwin provided a causal explanation for why our organs are disposed to function so as to promote survival, we still feel that the *value* of survival remains to be addressed. Also, most of us feel that normative values necessarily involve consciousness. If forced to consider what it would mean to have normative principles in the absence of consciousness or potential consciousness, could we formulate a coherent idea? Let's try to make the implications of this issue more vivid.

Being free to respond to normative principles—autonomy—seems to be an essential part of what we mean by thinking, cognition and related concepts. Let's suppose that Tavin, having a reasonably finely tuned philosophical acumen, were to consider Kim's argument. After thinking long and hard about Kim's premises—the causal closure of the physical, causal exclusion, the finding that nonreductive physicalism has unacceptable epiphenomenal implications, the irreducibility of qualia and the assumption that there are no good mentalistic alternatives—Tavin concludes that minimal physicalism is, indeed, near enough. Now let's suppose it were to occur to Tavin that a consequence of minimal physicalism is that the mechanistic collaboration of neurons firing in his brain has caused him to arrive at this conclusion, and would have done so *regardless of any normative considerations such as the soundness of the argument*. Wouldn't this tend to undermine Tavin's confidence in his conclusion, and in his ability to reason cogently in general? Nor, I think, would Kim be satisfied if the explanation for why we arrived at this conclusion (that minimal physicalism is near enough) had nothing to do with whether the conclusion was justified! He is appealing to us as rational, coherent persons; he wants us to adopt minimal physicalism based on normative principles of evidence and justification—and based on our ability to respond to these normative principles.

These are the kinds of issues that concerned Davidson. And although Kim rebutted Davidson's anomalous monism (successfully, I think) he never adequately addressed Davidson's motivation—the rationality, coherence and autonomy of our cognitive/intentional life. If these considerations are correct, then a significant part of our conception of mentality would be eliminated through the kind of functional reduction that Kim is advocating. Kim, of course, is aware of these objectionable consequences and remarks, "The best, or the most satisfying, outcome would have been the vindication of mental causation along the lines of nonreductive physicalism; that would have allowed us to retain mentality as something that is causally efficacious and yet autonomous vis-à-vis the physical domain."

(Kim, 2005, p. 159) Retaining mental autonomy would be best and most satisfying because much of our conception of what it means to be rational, coherent and, of course, scientific depends on it. Similar objections are frequently leveled against reduction on the grounds that most of our thinking about moral responsibility also depends on a robust conception of autonomy.

If Kim were to succeed in functionalizing cognitive/intentional properties in terms of their causal properties and then distribute normative evaluations over certain sequences of neurons firing (sequence N_1 is justified reasoning whereas sequence N_2 is not), it is not clear what the basis for this normative distribution would be. In functionalizing our cognitive/intentional properties Kim has redefined our concepts so that their normative content is no longer to be regarded as essential. We can therefore readily come up with possible worlds in which the normative distribution is much different than what we are used to or, indeed, in which there is no normative distribution at all. So Kim's complaint against the nonreductive physicalist applies equally well to his own minimal physicalism:

If you were to redistribute mental properties [substitute normative properties] over the events of this world any way you please—you might even remove them entirely from all events, making all of them purely physical—that would not alter, in the slightest way, the network of causal relations of this world; it would not add or subtract a single causal relation anywhere in the world!

We might call this *normative epiphenomenalism*, or perhaps even *explanatory epiphenomenalism*. Why *explanatory* epiphenomenalism? In his critique of an argument by Block and Stalnaker for type-identity reduction based on explanatory considerations, Kim shows that the transfer of causal effect between events does not automatically confer explanatory power. In the argument schema θ_1 , below, the laws of neurophysiology explain why N_1 causes N_2 . If we designate the body of neurophysiological laws as 'Neurophysiology' we can explain the relationship between N_1 and N_2 by showing how the statement of this relationship,

'Neural state N_1 causes neural state N_2 ,' follows from Neurophysiology. Since ' $M_1 = N_1$ ' and ' $M_2 = N_2$ ' permit substitutions, we can derive a causal relationship between M_1 and M_2 as follows:

(θ_1) Neurophysiology

Neural state N_1 causes neural state N_2

$M_1 = N_1$

$M_2 = N_2$

Therefore, M_1 causes M_2 .

Kim describes this reduction of M_1 to N_1 and M_2 to N_2 as the means by which we can save mental *causation* within a physicalist framework, although it does not save mental *explanation*. The identities $M_1 = N_1$ and $M_2 = N_2$ permit substitutions, but they do not add anything to the explanation, which is finished with line 2 in the derivation (θ_1). Based on this analysis, Kim makes it clear why reduction of the mental does not automatically save mental explanation: "...the only role the identities play is to enable us to restate the phenomenon that has already been explained.... All the explaining represented in the derivation occurs within neurophysiology....." (Kim, 2005, p.145)

From the fact that N_1 (along with Neurophysiology) explains N_2 , together with the fact that $M_1 = N_1$, it does not follow that M_1 explains N_2 , or that the occurrence of M_1 can *explain* anything. It is in this sense that Kim's reductive account may salvage mental causation and yet be accused of lapsing into not only normative epiphenomenalism, but explanatory epiphenomenalism as well. If functional reduction salvages mental causation, it saves it in name only.

Kim's Dilemma

In order for Kim to really save mental explanation he would have to show that the equalities in the argument schema, θ_1 , transmit explanatory force. What Kim requires is some sort of equality of meaning. Since it is not obvious that talk about beliefs and desires and the like are shorthand descriptions of neuronal interactions, he needs to make a convincing case that there is a plausible, even though non-obvious, sense in which $M_1 = N_1$ is analytic. More generally, by proposing that intentional/cognitive properties can be functionalized, Kim is

implicitly committed to the strong claim that the meaning of our cognitive/intentional concepts—including their responsiveness to normative considerations—can be analyzed in causal/functional terms. Based on our foregoing discussion, this claim is unsustainable because functionalization fails to capture our sense of the autonomy and normative character of our cognitive/intentional life.

This difficulty with trying to do justice to the normative character of our cognitive/intentional life is reminiscent of the problems faced by the program to “naturalize” ethics, as emphasized in G.E. Moore’s “open question” argument. Analogously, suppose the functionalist tells us that a sequence of cognitive/intentional states being cognitively/intentionally good (i.e. rational, coherent, logically justified, intelligible and so forth) just *means* that it contributes to the survival of the species. Then when the functionalist proves to us that a sequence of thoughts—say the evaluation of a body of evidence—has survival value in the requisite sort of way, that should be the end of the discussion. But surely it still makes sense to ask whether the evaluation of the evidence was justified. The point is that, in maintaining that we can functionalize cognitive/intentional properties and then reduce these functional counterparts to fundamental physics, the functionalist has “changed the subject,” to use Davidson’s expression. He has eliminated an essential part of our conception of intentional/cognitive properties—their autonomy and, in particular, their responsiveness to normative standards.

Kim is faced with a dilemma. He can (a) adopt an eliminativist attitude toward much of our conception of what cognition and intentionality involve, or (b) he has to make a convincing case that science will eventually make it clear why our cognitive/intentional concepts are functionalizable—and *analytically so*. If (a) he retreats into eliminativism he will face the usual charge that he is ignoring evidence. And, more serious, by eliminating autonomy and normative considerations from our conception of the mental, he would be wreaking havoc, not only with our conception of moral responsibility, but with

our conception of rationality and scientific objectivity, as well. An eliminativist position regarding the normative character of cognitive/intentional significance would degenerate into cognitive relativism since, by discarding normative standards for evaluating evidence and justification, there would be no difference between a good argument and gibberish. So, insofar as we take cognitive normative standards—standards of evidence, justification, rationality, coherence and the like—seriously, we cannot help but judge the eliminativist as implicitly contradicting himself by advancing a position for rational consideration that implies cognitive relativism.

Kim opts for (b), but is this option any better? Kim, himself, acknowledges that “...it is perhaps unlikely that we will have such [functionalist] definitions anytime soon.” (Kim, 2005, p. 165) Hilary Putnam is much stronger in his negative assessment of the prospects for functional reduction, and physicalism in general, characterizing them in terms of “science fiction.” In a sustained critique of reduction and physicalism—much of it devoted to earlier versions of Kim’s work—Putnam argues that, in light of current scientific understanding, claims for the reduction of mental concepts make no sense:

Saying “Science may someday find a way to reduce consciousness (or reference, or whatever) to physics” is, *here and now*, saying that science may someday do we know-not-what we-know-not-how. And from the fact that those words may in the *future* come to have a sense we will understand, it no more follows that they *now* express anything we can understand than it follows from the fact that I may *someday* learn to play the violin that I can now play the violin.

In particular, I have argued that neither in the case of appearances (including looks) nor in the case of the propositional attitudes do reductionist claims make real scientific or philosophical sense.” (Putnam, 1999, p. 173)

Kim’s current position concedes that phenomenal consciousness and appearances (which he understands in terms of qualia) cannot be reduced. That narrows the argument to cognitive/intentional properties

(note Putnam's allusion to reference and propositional attitudes).

Putnam's criticism can be improved upon, I think, and in its present form it can be objected to on the grounds that philosophy has the right, and the obligation, to try map out the logical terrain ahead of where science currently is. If philosophy is reluctant to speculate on what might turn out to provide fruitful directions for scientific research, others will do it anyway, and without the benefit of the standards of precision that philosophy has developed over millennia.

Nevertheless, I think Putnam is right that there is a conceptual incoherence with the program to functionally reduce the cognitive/intentional domain. Part of it has to do with the paradox of arguing for *eliminativism*, while expecting your argument to be rationally convincing. There is a similar paradox for those who, like Kim, believe that functional reduction will *conserve* what is essential about our conception of the cognitive/intentional domain.

As noted above, materialists traditionally regarded functional reduction as supporting the elimination of cognitive/intentional attributions to a system. Thus the rejection of vitalism or any kind of purposive explanation of the organization of biological systems in the 19th century was justified by the claim that once a reductive, causal explanation of a function had been provided, any attribution of cognitive/intentional properties would be superfluous. Based on classical mechanics, natural selection, molecular biology and the like, materialists believed they could eliminate any trace of cognitive/intentional properties from the fundamental canon of physical theory. Explanations of natural phenomena in terms of causal mechanisms are taken to *exclude* explanations in terms of Divine purposes, entelechies, vital forces or anything involving attributions of cognitive/intentional properties to the physical domain. Contemporary physicalism, of course, claims this intellectual inheritance. Thus it is not surprising that, in the current debates, physicalists tend to have little patience for the idea that the observer in relativity theory, or measurement in

quantum mechanics, will require the introduction into physics of anything involving consciousness or cognition. In this context, if the functional reduction of cognitive/intentional properties to neuroscience is supposed to be conservative, rather than eliminative, then this would create an implicit contradiction. Cognitive/intentional properties are not conserved by being functionalized if those functional properties are in turn reduced to fundamental physics which is allowed only causal properties to the exclusion of cognitive/intentional ones.

Hence, whether the functionalist intends his reduction to be eliminative or conservative of cognitive/intentional properties, he is implicitly contradicting himself. If eliminative, the contradiction comes in because eliminativism implies cognitive relativism and, yet, the functionalist believes his argument is based on good reasons. If conservative, as for Kim, the contradiction comes in because the proposed reduction is to a reductive base that is devoid of cognitive/intentional significance.

So the conceptual problem for functional reduction is not based on its outrunning contemporary neuroscience, as Putnam seems to suggest. There is a conceptual incoherence inherent in functional reduction because, either the functionalization step is intended to be eliminative, or the reduction to fundamental physics turns out to be eliminative anyway, of what is essentially cognitive and intentional about the mental domain. These considerations, together with the difficulties besetting reductionist alternatives to functional reduction—bridge-law reduction and identity reduction founder on the problem of multiple realization—lead to the conclusion that reductionism itself is problematic.

So, where are we in the argument? Here is Kim's assessment of what is required for mental causation:

What has become increasingly clear after three decades of debate is that if we want robust mental causation, we had better be prepared to take reductionism seriously, whether we like it or not. But even if you are ready for reductionism, it doesn't

necessarily mean that you can have it. For reductionism may not be true (Kim, 2005, p.22).

Kim's cautious note in the second two sentences turns out to be well-founded, as I believe I have shown. His urgency in endorsing reduction in the first sentence overstates his case and I would prefer to say that, if we want robust mental causation—where it is mental in more than name only—we had better be prepared to take interactionism seriously, whether we like it or not.

Jerry Fodor famously objected that, "... if it isn't literally true that my wanting is causally responsible for my reaching ... and my believing is causally responsible for my saying ... then practically everything I believe about anything is false and it's the end of the world." (Fodor, 1990, p.156) Fodor's eloquent appeal to common sense cannot be assuaged with a truncated conception of mental causation. If my argument has merit, then an adequate account of mental causation will require a robust conception of cognitive autonomy, and this implies interactionism.

Acknowledgments

I would like to thank Stephen Barker, Joel Marks, Jonathan Shear, and Ed Kelly for valuable feedback on earlier presentations of this material; and my colleagues and friends at Maharishi University of Management, especially Fred Travis, Alarik Arenander, Robert Boyer, Robert Klauber, Russell Hebert, Robert Oates and, above all, John Hagelin. Also, I would like to thank Deborah Raub and Rolf Erickson for their improvements to both the style and substance, and I am grateful for the continuous dialogue with my brother Peter Scharf and my wife, and intellectual collaborator, Susan Scharf.

References

- Davidson D. Mental events. Originally in Foster L and Swanson J (eds). *Experience and Theory*. University of Massachusetts Press, 1970. Reprinted in (Davidson, 1980).
- Davidson D. *Essays on Actions and Events*. New York and Oxford: Oxford University Press, 1980.
- Dempsey L and Shani I. Dynamical agents: Consciousness, causation, and two specters of epiphenomenalism. *Phenomenology and the Cognitive Sciences: ISSN 1568-7759 (Print) 1572-8676 (Online)*, 2008.
- Feigl H. The 'mental' and the 'physical'. Originally published in Feigl H, Scriven M and Maxwell G. (eds). *Minnesota Studies in the Philosophy of Science*, vol. 2. Minneapolis: University of Minnesota Press, 1958.
- Fodor J. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press, 1990.
- James W. *The Principles of Psychology*. New York: Henry Holt & Company, 1890. Available online at <http://psychclassics.yorku.ca/James/Principles/index.htm>.
- Kant I. *Groundwork of the Metaphysic of Morals*. Translated by Paton HJ. New York: Harper and Row, 1964. Originally published by Kant in 1785.
- Kim J. *Supervenience and Mind*. Cambridge: Cambridge University Press, 1993.
- Kim J. *Mind in a Physical World*. Cambridge, MA: MIT Press, 1998.
- Kim J. *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press, 2005.
- Kim J. *Philosophy of Mind (2nd ed)*. Cambridge, MA: Westview Press, 2006.
- Nagel T. What is it like to be a bat? *Philosophical Review* 83:435-450, 1974.
- Putnam H. *The Threefold Cord: Mind, Body, and World*. New York: Columbia University Press, 1999.
- Scharf D. Quantum measurement and the program for the unity of science. *Philosophy of Science* 1989; 56:601-623.