

Why Jaegwon Kim's Physicalism is Not Near Enough

An Implicit Argument for a non-Cartesian Interactionism

Part II

David Scharf

Abstract

In Part I we critiqued Kim's thesis that cognitive/intentional properties can be accounted for in terms of functional reduction to neurophysiology. We saw that this thesis is conceptually incoherent and implicates Kim in *eliminative* materialism, despite his sanguine belief that he is *saving* what is distinctive about mind by providing for mental causation. These considerations argue for the autonomy of at least some aspects of the cognitive/intentional domain. Here, in Part II, we will further critique minimal physicalism by showing that Kim's radical demarcation between phenomenal consciousness and cognitive/intentional properties is untenable. Contemporary research in cognitive neuroscience documents the principle, anticipated by William James, that consciousness is functional and 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. Although there remain compelling reasons to reject a Cartesian model of radically distinct substances, both threads of this critique (Part I and Part II) indicate the need to take seriously a more sophisticated interactionism, where consciousness is metaphysically fundamental and interacts with biophysical processes in the brain. In order to avoid the extreme conclusion that consciousness is absolutely irrelevant to our cognitive/intentional functioning, Kim proposes that, although individual qualia are epiphenomenal, qualia relations may have cognitive/intentional significance. However, I show that his proposal involves a paradox of conflicting dependencies, which may be characterized as supervenience collision.

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

NeuroQuantology 2010; 4: 614-625

Corresponding author: David Scharf

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 25, 2010.

ISSN 1303 5150

www.neuroquantology.com

Consciousness and Cognition—Are qualia really free of cognitive/intentional significance?

Now, let's turn our attention to Kim's radical partition between phenomenal consciousness and cognitive/intentional properties. The explanatory gap, or hard problem of consciousness, is based on the supposition that, once we have mapped out all the minute subregions of the brain and documented all their neurological, causal processes, we will not have explained the fact of subjective experience, with its related characteristics of privileged access, first-person perspective and sense of self. If my foregoing argument in Part I has been correct, then we will not have explained the rationality and coherence of the mental domain either. But even this description of the problem grants too much to the reductionist, since it begs the question of whether a complete description of neurological processes is possible *without* understanding the constraints that rationality, coherence and autonomy may place on neurological processing. If interactionism is correct, a completed neurological account would have to take into account cognitive/intentional intervention, where this is understood as involving an autonomous source of agency.

I have suggested that this may involve fundamental physics in a way not anticipated in Kim's essentially classical worldview. Indeed, there is a burgeoning field of quantum neuroscience based on just this premise.

Here, in Part II, I will make the case that this autonomous source of agency is consciousness itself, consistent with William James's viewpoint, argued more than a century ago:

The particulars of the distribution of consciousness, so far as we know them, point to its being efficacious (James, 1890, p.138).

Kim appears to come part of the way toward acknowledging this same point, remarking that, "Consciousness in humans may be a way of monitoring certain internal goings-on; that may be its biological function. But, that is not to say that the essence of consciousness is such monitoring," he continues, "or that whenever

there is such monitoring, there must be consciousness" (Kim, 2006, p.220). This suggests that there are two important aspects or properties of consciousness, its phenomenology and its functionality.

In what follows I will argue, against Kim, that these two properties of consciousness are essentially entangled. If the explanatory gap implies the autonomy and irreducibility of phenomenal consciousness (which Kim is prepared to admit with some qualifications to be explained below), it follows that unless he can successfully disentangle functional consciousness from phenomenal consciousness he should also acknowledge the irreducibility of at least this much functionality. In other words, he ought to admit some form of interactionism.

If these considerations prove to be correct, we will have established a second argument for interactionism, in addition to the autonomy argument of Part I. Moreover, if there is an essential entanglement between phenomenal consciousness and functional consciousness, then the two arguments are interrelated. To anticipate the arguments to be presented below, assume we successfully show here, in Part II that—in agreement with William James—explanatory gap considerations support interactionism. The explanatory gap points to the irreducibility of phenomenal consciousness, and if there is functionality which is intrinsic to phenomenal consciousness, then this functionality is also irreducible. Then my conclusion will suggest that the autonomy of the cognitive/intentional domain, as discussed in Part I, is grounded in the autonomy of consciousness itself.

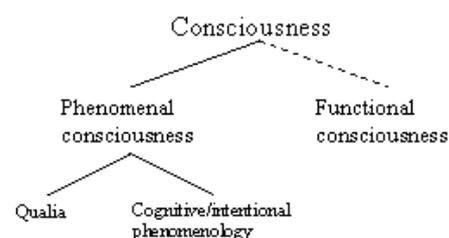


Figure 1. Aspects of Consciousness. Is phenomenal consciousness functional?

Cognitive/Intentional Phenomenology

Phenomenal consciousness is often characterized as “*what it is like*” to be having such and such a subjective experience. Qualitative subjective experiences (qualia) include sensory qualities like colors, sounds, and so on, as well as physical sensations like pain and pleasure. Whether occurrent beliefs and desires, and generally cognitive/intentional states that we are conscious of having, should be considered qualia seems to be a semantic debate that we need not resolve here. Certainly we can be phenomenally conscious of our cognitive/intentional states. Galen Strawson highlights the existence of *understanding-experience* by considering the difference when a Frenchman, Jacques, and an Englishman, Jack, both listen to the news in French, but only Jacques speaks French (Strawson, 1994). There is a difference in what it is like in that only Jacques is having what Strawson characterizes as an understanding-experience of the news. Horgan and Tienson refer to this as *cognitive phenomenology* (Horgan and Tienson, 2002) and, in keeping with Kim's terminology; I will refer to it as cognitive/intentional phenomenology. It seems uncontroversial that phenomenal consciousness includes both qualia and cognitive/intentional phenomenology. The controversial question is whether phenomenal consciousness is functional. In Figure 2 these two aspects of consciousness—phenomenal consciousness and functional consciousness—are shown as if distinct, but the actual nature of the relationship between these aspects remains to be clarified. The line to “Functional consciousness” is dashed to indicate that functional consciousness is depicted as separate for heuristic purposes only.

Having cognitive/intentional significance implies functionality, so if phenomenal consciousness is cognitive/intentional, it will be functional. Evocative of Edmund Husserl, in a recent article Horgan and Kriegel introduce the term “phenomenal intentionality” in order to specifically acknowledge the intrinsic cognitive/intentional significance of phenomenal consciousness:

...there exists phenomenal intentionality.
The phrase “phenomenal intentionality”

denotes a kind of intentionality that phenomenally conscious states exhibit, and moreover exhibit precisely *in virtue* of being phenomenally conscious states, that is, in virtue of their specific phenomenal character. The first thesis, then, is that there is such an intentionality; that there exist phenomenally conscious states that are intentional and intentional in virtue of their phenomenal character. More precisely, the claim is that there are mental states that instantiate properties whose nature is both phenomenal and intentional: the constitutive phenomenal character of such properties is intrinsically intentional” (2008, section 2).

They seem to intend the strong thesis that *all* phenomenal consciousness has intrinsic cognitive/intentional significance, and I regard this as a plausible thesis. Even the humble color patch seems to have intrinsic cognitive/intentional significance as evident, for example, in the warmth of redness and the coolness of green. Moreover, as noted by George Berkeley, it seems impossible to experience color without a sense of extendedness and shape, however indefinite that shape may be. The geometrical properties seem to be intrinsic to the experience of color—that is, the color quale seems to have intrinsic complexity, and awareness of complexity would seem to be cognitive/intentional. As against these intuitions, Kim must regard qualia as simple and without intrinsic cognitive/intentional significance, as we will see shortly.

Whether or not Horgan and Kriegel are committed to regarding all subjective experience as intrinsically cognitive/intentional, they unambiguously affirm that “... phenomenal intentionality is exhibited not only by sensory states, but also by more cognitive, ‘intellectual’ states” (2008, section 2). In contrast, Kim not only does not accept that sensory states can have intrinsic cognitive/intentional significance but his theoretical commitments require that he deny intrinsic phenomenology to cognitive/intentional states. He needs to tease apart phenomenal experience from cognitive/intentional significance in order to account for the twin obligations of phenomenal irreducibility (explanatory gap) and mental causation (via the functional reduction of cognitive/intentional properties) without resorting to interactionism. His proposal is

to radically partition qualia (which will be the sole repository for phenomenal experience) from the cognitive/intentional domain (which will therefore have to be regarded as intrinsically non-experiential). But then how is he going to handle cognitive/intentional phenomenology? Kim has an idea for a solution, but I don't think it works.

The difficulty for Kim is to explain what it means for consciousness to be functional, if phenomenal consciousness—or some aspect of it—is autonomous and irreducible. An obvious solution is interactionism, but Kim won't go there. Well, then, how can phenomenal consciousness have cognitive/intentional significance, given that cognitive/intentional significance implies functionality?² But before examining Kim's proposed solution in detail, it will be useful to look at the evidence from

² Horgan and Kriegel propose that "... phenomenal intentionality is the only kind of non-derivative intentionality" (Horgan and Kriegel, 2008, section 2). This would imply that our unconscious mental life—at least insofar as it is really unconscious—is cognitive/intentional in a secondary or derivative sense, like expressed language or even like street-signs or flags. Incidentally, Kim, in at least one discussion, seems to associate phenomenal consciousness exclusively with qualitative sensory consciousness. Thus, he remarks that:

In general, mental occurrences that we call "experiences" appear to be those that possess phenomenal properties. Sensing and perceiving are experiences, but we do not think of believing and thinking as experiences. If this is so, the idea of phenomenal character and the idea of there being something it is like may come apart, though only slightly. For it certainly seems that there is something it is like to believe something, to suspend judgment about something, to wonder about something, or to hope for something. But as we saw, at least many instances of these states do not seem to have any phenomenal character (Kim, 2006, p.210).

It seems to me that most people *do* think of conscious believing and thinking as experiences. So I am classifying occurrent cognitive/intentional states as belonging to phenomenal consciousness, even if they are not sensory. Moreover, there is an obvious tension between Kim's acknowledgement that, "...it certainly seems that there is something it is like to believe something, to suspend judgment about something, to wonder about something, or to hope for something," and his denial that these constitute experiences. As I will try to show in some detail, this tension is an indication of the untenability of his position.

contemporary cognitive neuroscience for the functionality of consciousness.

Global Workspace Theory—the evidence from cognitive neuroscience

Contrastive analysis is Bernard Baars' term for the comparative study of conscious processes and cognitively similar, unconscious processes. Examples include unconscious perception, subliminal priming and automaticity (Baars, 1997 and 2002).

Unconscious perception has been studied extensively over the past few decades and occurs, for example, when words or pictures are either masked so as to be unrecognizable or are flashed before a subject too quickly to register consciously. More recently binocular rivalry has enabled precise studies of conscious versus unconscious perception. Binocular rivalry occurs when the two eyes are presented with different and incongruous images. The subject's awareness typically switches between the images, so that only one is conscious at any given time, although both the conscious and the unconscious percept activate some of same parts of the visual cortex up to a point. In general, conscious perception, as opposed to unconscious perception, activates a wide range of brain resources in addition to those dedicated to visual analysis.

Subliminal priming can be quantitatively measured in word-recognition tasks. For example, words presented too quickly to identify consciously still have cognitively measurable effects, in that they influence subsequent word recognition. Nevertheless, the cognitive effects of subliminal priming, like those of unconscious perception generally, are quite limited compared to the global access provided by conscious attention.

Automaticity also shows that, although there are often cognitive similarities between conscious and unconscious processes, there are also important differences that point to the functional and adaptive role of consciousness. As we practice a skill, it becomes automatic and no longer requires conscious attention. Nevertheless, when confronted with a new, unexpected or critical situation, our repertoire provides no substitute for conscious attention.

The evidence shows unambiguously that, as William James anticipated, consciousness is functional. There are measurable neurological and behavioral differences between conscious processes and cognitively similar unconscious processes and, moreover, these differences indicate the adaptive role of consciousness. Despite the limited capacity of conscious attention, it plays a distinctive role in facilitating the integrated functioning of distributed neurological components, the global mobilization of cognitive resources, as well as enabling robust learning. On the role of consciousness in learning, Baars remarks, "At an everyday level this simply comes down to the obvious point that in order to learn something, we have to pay attention to it. But of course, 'paying attention' means that we stay conscious of something." (1997, p. 162).

Global workspace theory is based on a theater metaphor, where the stage represents the integrative role of working memory, and where the essential function of consciousness in working memory is represented by the spotlight on the stage. The spotlight illuminates the action, making it widely accessible to the whole audience, where the audience represents the distributed, cognitive resources of the brain. The spotlight metaphor for consciousness highlights both its limited capacity and its importance for making significant information widely accessible to the brain's distributed resources. Global workspace theory, then, is intended to characterize the function of consciousness in providing global access to new, unexpected or critical information.

The contemporary work in cognitive neuroscience documents the point William James expressed so elegantly:

It is very generally admitted, though the point would be hard to prove, that consciousness grows the more complex and intense the higher we rise in the animal kingdom. That of a man must exceed that of an oyster. From this point of view it seems an organ, superadded to the other organs which maintain the animal in the struggle for existence; and the presumption of course is that it helps him in some way in the struggle, just as they do. But it cannot help him without being in

some way efficacious and influencing the course of his bodily history (James, 1890, p.138).

For James, there is no fundamental distinction between functional consciousness and phenomenal consciousness; they are just two aspects of the same phenomenon. When he is speaking of consciousness, he means phenomenal consciousness and he is making the plausible claim that phenomenal consciousness is functional and adaptive.

However, it is possible to conceptually differentiate phenomenal consciousness from functionality (let's not say "from functional consciousness," because how would it be consciousness without phenomenology?). Since the primary behavioral indicator (in humans) for consciousness is verbal reports, it is possible to acknowledge the functionality of what we call "consciousness," with the attendant verbal reporting and other behavioral cues, but still deny phenomenal experience. In other words, the suggestion would be that some of the subjects in our cognitive experiments are zombies, neurologically and behaviorally like us but without phenomenal awareness. But it's hard to take the zombie hypothesis too seriously, even as a conceptual possibility, as Kim himself acknowledges:

In fact, I believe the zombie hypothesis is untenable To hold onto the zombie hypothesis, we must apply a massive "error theory" to these creatures—namely that all their (positive) phenomenal assertions are false. I believe this is incoherent. We must grant that the creatures have inner consciousness, although the qualitative character of their consciousness remains undetermined (2005, p.169 fn).

By maintaining that "the qualitative character of their consciousness remains undetermined," Kim is holding open the possibility of qualia inversion. But first there is an important logical point.

Physicalists — Kim included — sometimes speak as though establishing functionality is all that is needed to establish reducibility, ignoring the possibility of interactionism. If the explanatory gap is evidence that phenomenal consciousness is irreducible, and if evidence from cognitive neuroscience can be taken to indicate that phenomenal consciousness is functional,

these two premises together support an interactionist conclusion, which is the conclusion that William James drew. By conceding that “the zombie hypothesis is untenable,” Kim also is implicitly acknowledging that functional irrelevance is not necessary for irreducibility. This is so because, if phenomenal consciousness were functionally irrelevant and epiphenomenal, then its absence could make no difference to the coherence of any error theory—it could have no cognitive/intentional significance for anyone, including us as detached observers debating the issue. Epiphenomenalism and eliminativism are, in all material respects, indistinguishable. So, if positive verbal reports are, indeed, responsive to the presence of phenomenal consciousness—as the incoherence of a massive error theory would indicate—then phenomenal consciousness is functional, even though the explanatory gap points to its irreducibility.

If these considerations are correct, Kim has already made a major concession to interactionism by conceding the incoherence of the zombie hypothesis. And his position will become increasingly untenable as we get further into the details.

The Relationship between Qualia and Cognitive/Intentional Phenomenology

Although Kim recognizes the incoherence of the zombie hypothesis—thus implicitly acknowledging that the presence or absence of phenomenal consciousness has cognitive/intentional significance—he is willing to countenance the possibility of inverted spectra. When we drive towards a traffic light and it's green, we go and, if it's red, we stop. According to the inverted spectrum scenario, it's possible that some of us could have our qualia systematically reversed without its making any practical difference, so that what we experience as green corresponds to what others see as red and vice versa.

But what about cognitive/intentional phenomenology, what Galen Strawson characterizes as understanding-experience? What if, by analogy to qualia inversion, phenomenal inversions were possible for the more sophisticated cognitive/intentional phenomenology?

In Part I, we considered an argument schema, θ_1 , similar to one which Kim borrowed from Block and Stalnaker, which he used to show that the identification of mental properties with physical properties may provide for mental causation, while failing to salvage mental explanation. I extended this reasoning to show that Kim's functional reduction leads to an unpalatable explanatory epiphenomenalism regarding cognitive/intentional properties. The incoherence inherent in explanatory epiphenomenalism will be more vivid if Kim were to allow the possibility of phenomenal inversions for understanding-experiences.

Thus, suppose that Tavin has an understanding-experience Q_1 , while thinking about the evidentiary force of considerations of causation, contiguity and the causal closure of the physical. Let us assume that these are the considerations that Kim believes should convince us that his minimal physicalism is near enough. Q_1 is the subjective experience one would ordinarily have when thinking about these premises. Moreover, Q_1 occurs in conjunction with M_1 , which embodies the intentional/cognitive properties usually associated with Q_1 , and $M_1 = N_1$, where N_1 is the neural basis for M_1 . Thus,

$$Q_1 \leftarrow M_1 = N_1$$

where ' \leftarrow ' indicates that Q_1 is an epiphenomenon and '=' means that M_1 and N_1 are the same event. (M_1 describes the event in intentional/cognitive terms whereas N_1 gives its neurophysiological description.)

Now, as a result of thinking about causation, contiguity and the causal closure of the physical, Tavin comes to the conclusion that: Yes, Kim's physicalism is near enough. We can represent this conclusion by

$$Q_2 \leftarrow M_2 = N_2$$

where ' $Q_2 \leftarrow M_2$ ' is the mental state consisting of the subjective awareness (Q_2) associated with Tavin's conclusion together with the corresponding intentional/cognitive properties (M_2). N_2 is the neural basis for this mental state.

In the argument schema θ_2 , below, as in the original schema, θ_1 , the laws of neurophysiology explain why N_1 causes N_2 .

The following continuation of the analysis sharpens the problem for Kim of rendering phenomenology causally irrelevant.

(θ_2) Neurophysiology

Neural state N1 causes neural state N2

$Q1 \leftarrow M1 = N1$

$Q2 \leftarrow M2 = N2$

Therefore, M1 causes M2 and Q2.

Q1, the subjective experience of thinking about causation, contiguity and the causal closure of the physical, has neither causal efficacy nor explanatory value. Hence, Kim's analysis permits the following kind of inverted-phenomenology scenario. When Tavin has neural state N1, by Kim's reductive thesis ($M1 = N1$), he *ipso facto* is "thinking about causation, contiguity and the causal closure of the physical" even though—due to some sort of phenomenological inversion—he is having the understanding-experience that Jacques would ordinarily have when listening to the news in French. If we call this understanding-experience $Q1'$, then $Q1' \leftarrow M1 = N1$ can be substituted for line 3 of derivation (θ_2) without altering the result in any way! We can still conclude that

Therefore, M1 causes M2 and Q2.

In other words, his "thinking about causation, contiguity and the causal closure of the physical" (M1) caused him to conclude that Kim's physicalism is near enough (M2 and Q2) even though his subjective experience ($Q1'$) was that of listening to the news in French. This would involve a massive error theory on the same order as what would be needed to accommodate the zombie hypothesis. But how can Kim permit qualia inversion without thereby rendering cognitive/intentional phenomenology invertible as well?

Kim's proposal is ingenious, even if not ultimately successful. Here is a statement of what he has in mind:

Intrinsic qualities of qualia are not functionalizable and therefore are irreducible, and hence causally impotent. They stay outside the physical domain, but they make no causal difference and we won't miss them. In contrast, certain important relational facts about qualia, in particular their similarities and

differences, are detectable and functionalizable, and can enjoy causal powers as full members of the physical world. (2005, p.173)

The first thing to notice is: while Kim is right that, if intrinsic qualities of qualia are not functionalizable they are irreducible, the converse is questionable. If we can show, in opposition to Kim, that intrinsic qualities of qualia are functionalizable, it does not follow that they are reducible. Interactionism remains a viable alternative. The second is that Kim places all the cognitive/intentional weight on "relational facts about qualia, in particular their similarities and differences...." But what would be the basis for qualia comparison if the qualia themselves make no causal difference and are devoid of intrinsic intentional/cognitive significance? Kim himself worries about "...the question whether it is possible to combine qualia epiphenomenalism with full causal efficacy of qualia similarities and differences" (2005, p.174).

Let's try to make this worry concrete.

The Problem of Supervenience Collision

Kim's proposed explanation of cognitive/intentional phenomenology denies that individual qualia have any intrinsic cognitive/intentional content. But while individual qualia have no cognitive significance on their own, according to Kim, qualia comparisons *do* have cognitive significance.

Therefore it becomes vitally important for him that we be able to sharply individuate and distinguish qualia. This, of course, places him at odds with the holistic theme of contemporary epistemology. According to the contemporary way of thinking, although there are vague, pragmatic distinctions between subjective experience and intentional/cognitive significance (these distinctions may be useful in one context or another), when we try to locate an absolute demarcation, more questions are raised than can be answered.

For example, of course we sometimes distinguish between an experience and its cognitive significance. We may see a pair of shoes at the door and remark, "Oh, this means that Tavin is at home." But even if we

are here distinguishing *this* meaning from the perception, it does not follow that the perception is free of cognitive content altogether. Obviously, we are seeing the shoes *as* shoes, for example, as well as evidence that Tavin is at home. But Kim will respond (as H. H. Price responded) that we must distinguish *seeing* from *seeing as*. But is there really anything that we just see, without intrinsic intentional/cognitive significance? We don't need to rehearse the arguments about the theory-ladenness of perception to see that Kim is fighting an uphill battle by maintaining that sensory experiences are qualitatively simple and absolutely free of intrinsic cognitive/intentional significance. Rather, where we draw the line between qualia and intentional/cognitive properties plausibly seems to be flexible and relative—a pragmatic distinction dependent on the circumstances—but Kim's argument requires an absolute, ontological partition between perception and cognitive significance.

Aside from these epistemological considerations, there is a logical problem with Kim's handling of qualia and their relations: the problem of *supervenience collision*. This problem emerges from the combination of two premises, both of which are implicit in Kim's position. The first premise is that qualia comparisons supervene on the qualia themselves; whether or not a set of qualia are all the same depends on the qualia.

- (1) Whether or not qualia q_1, \dots, q_n are the same is supervenient on q_1, \dots, q_n .

Let's call this *qualia supervenience*. If q_1 is a red quale for Tavin, q_2 is a green quale for Tavin and $R(q_1, q_2)$ means that q_1 is different than q_2 , it would follow that if other instances of q_1 and q_2 presented themselves to Tavin they would again be different.³ The sameness/difference relationship between

³ Kim is not committed to saying that the difference between q_1 and q_2 necessarily means that Tavin will perceive them as different. But if Kim denies that their difference/sameness has any *relevance* for how Tavin perceives them, then he could not defend the thesis that their comparison has cognitive/intentional significance.

qualia is dependent on the underlying qualia and their intrinsic properties.

Now Kim is also committed to some qualia comparisons, especially differences and similarities, having cognitive/intentional significance. That is, these relationships "... are behaviorally detectable, and this opens a way for their behavioral functionalization" (2005, p. 172). Ignoring the possibility of interactionism, for the present, functionalization paves the way for physical/biological reduction. So Kim is committed to:

- (2) Sameness or difference between qualia can have cognitive/intentional significance; in these instances the sameness or difference relationship is functionalizable and supervenient on physical/biological processes.

Combining (1) and (2) gives us:

- (3) Some qualia relationships R are cognitive/intentional properties, which are supervenient on qualia *and* supervenient on physical/biological processes.

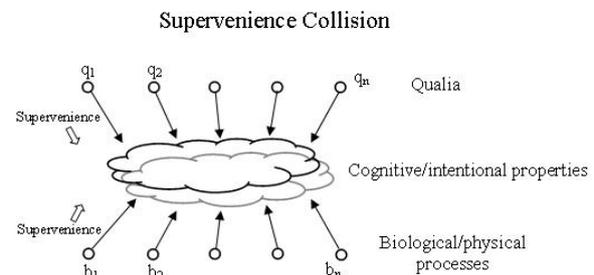


Figure 3. Supervenience collision: Cognitive/intentional properties with potentially conflicting dependencies.

But this involves supervenience collision (see Figure 3), because qualia are an epiphenomenal mental residue, with no cognitive/intentional significance. This mental residue "...has no place in the causal structure of the world and no role in its evolution and development" (2005, p.171). The epiphenomenalism of qualitative content means—if it means anything at all!—that we should be able to reassign qualitative content to a quale without disturbing anything in the physical domain. But let's consider Tavin's qualia, q_1 and q_2 , again, where q_1 is Tavin's experience of red and q_2

is his experience of green. Now if we altered q_1 a little, to pink, say, it shouldn't make any difference to anything. It's an epiphenomenon, after all. Tavin would still be stopping at traffic lights when he experiences q_1 . But now let's alter q_1 a little more, this time to *green*. The cognitive/intentional significance of $R(q_1, q_2)$ would suddenly change to signifying that q_1 is now the *same as* q_2 , resulting in erratic behavior at traffic lights. This would generate a contradiction, since, by hypothesis, we haven't changed any biological/physical processes. If these considerations are right, then Kim can't make qualia contents epiphenomenal if qualia comparisons have cognitive/intentional significance.

Evidently, Kim's effort to distinguish the supposed cognitive non-significance of individual qualia, from the cognitive significance of qualia comparisons, is unworkable. And if he is thereby obliged to allow that our neurophysiology is responsive to, or influenced in any way by, qualia and their intrinsic qualities, he's going to lose causal closure of the physical. But so what? Maybe it's time to put interactionism back on the table for serious consideration.

Conclusion

As we've seen, the arguments from Part I and Part II both support interactionism. The argument from Part I emphasized that the rationality and coherence of the cognitive/intentional domain (under favorable circumstances) reflects our ability to respond to normative considerations. This, in turn, implies a robust cognitive/intentional autonomy and the independence of the cognitive/intentional domain from determinative physical causes. In Part II, we challenged Kim's effort to separate out the functionality that is evidently inherent in consciousness. Moreover, given that Kim acknowledges that explanatory-gap considerations point to the autonomy of consciousness, its functionality (by intervention in neurophysiological processes) would violate what, in the Introduction, we called Type A causal closure of the physical. These two arguments are interrelated if, as seems likely, the autonomy inherent in cognitive/intentional significance

is in fact the same autonomy that explanatory gap considerations point to for consciousness. If this is right, consciousness is metaphysically fundamental, it has inherent cognitive/intentional significance and it interacts with the physical processes of the brain and nervous system. And, after all, isn't this just what common sense tells us?

But contemporary philosophers have been reluctant to draw an interactionist conclusion, primarily because of considerations of causation and contiguity. If we drink a cup of coffee, it affects our conscious experience by influencing the functioning of our brain. And, although our beliefs and desires seem to likewise affect our behavior, the proximate causes of our behavior are found in the neurophysiological processes in the brain and nervous system. These considerations, together with a not-sufficiently-critical confidence in (Type A) causal closure of the physical, have led the mainstream philosophical community to believe that some sort of physicalism just *has* to be right, even if this means overriding common sense. So interactionism will not be the first choice of many philosophers, even those who may agree that Kim's minimal physicalism is flawed. Their first choice: to reconsider nonreductive physicalism.

But our discussion in Part I determined that Kim's critique of nonreductive physicalism is fundamentally correct: it will resign the mental domain to an essentially epiphenomenal status. The attraction of nonreductive identity theories or dual-access theories is precisely that they promise to preserve mental causation. But, like Kim's functional reduction of the cognitive/intentional domain, these theories embed mental/neurophysiological processes in a broader nexus of physical causes for which the normative significance of the piggy-backing mental phenomena (whether understood in terms of mental properties, access or descriptions) is causally irrelevant. A persuasive theory must preserve what is distinctive about our cognitive/intentional life—its rationality, coherence and ability to respond to normative constraints. These considerations demand an explanation which gives priority to the cognitive/intentional domain itself. Moreover, although there are secondary

senses of cognitive/intentional significance (street signs, flags, characters on a page and the like), in its primary meaning cognitive/intentional significance is essentially embedded in consciousness.⁴

To summarize, our critical analysis of physicalism has resulted in several suggestions that may provide the outline of a comprehensive alternative to the physicalist perspective in the philosophy of mind. First, the unworkability of Kim's demarcation between consciousness and cognitive/intentional properties suggests that consciousness may have intrinsic cognitive/intentional significance. In fact, although Kim framed this discussion in terms of qualia, I think this may well have broader implications. What I have in mind is the enduring theme in ethical theory of our mind as incorporating input from our higher Self or higher aspects of our consciousness, whose intrinsic intelligence and purposes compete in our nervous system with our biological drives. Second, and also consistent with this theme, is the recognition that our cognitive/intentional life has—at least when we are at our best—a robust autonomy which supports our ability to respond to normative considerations. This autonomy is essential to our conception of ourselves as rational and morally responsible. Third, consciousness itself can be causally efficacious, and while this efficacy may be transmitted by the biophysics of our brain and nervous system, it refuses to reduce to biophysics.

Together these suggestions point to interactionism, but exactly what *kind* of interactionism? In the seventeenth century, when modern science was in its infancy, it may have seemed reasonable for Descartes to propose an interactionism between mind and matter as radically distinct substances, where matter was the subject of the new mathematical physics, and mind was outside its domain.⁵ But today Cartesian

interactionism seems decidedly less plausible because, among other reasons, if something—even if it is consciousness itself—is interacting with the physical domain, physics is going to want to understand it, incorporate it into the scientific domain, and model its interactions with suitable equations.⁶ As explained in the Introduction, Type B causal closure is practically tautologous, and it follows from this that any interaction between mind and matter can be modeled within physics—it's just that the domain of physics will not be exclusively physical (in the sense of physical,) and perhaps not physical at all, except at the more surface levels which can be understood in terms of classical physics.⁷

So, rather than an interaction between radically distinct, Cartesian substances, it would make more sense, nowadays, to suppose that mind and matter are two expressions of a common underlying substance and, what's more, this theme is in keeping with the trend toward unification that has predominated in contemporary science—particularly, in recent advances in theoretical physics. In Western thought there have been two main representations of this theme: neutral monism and idealism.

The neutral monism of Ernst Mach, William James, Bertrand Russell and others at the turn of the 19th century has recently been enjoying something of a revival, drawing serious consideration from leading figures in consciousness studies, including David Chalmers and Stuart Hameroff. But the considerations we have explored in these pages suggest that it is *consciousness* that is autonomous and primary—if consciousness were dependent on an abstract, neutral substance it would be neither. And, moreover, nothing in our previous discussion motivates a parallel requirement for the physical domain to be regarded as autonomous and primary. This points toward a monistic view in which the primary stuff of reality is consciousness; this view also has an influential and distinguished

⁴ Horgan and Kriegel (2008, section 3) emphasize that "phenomenally intentional states are the only states that are uncontroversially, unquestionably, paradigmatically, prototypically mental."

⁵ It should be noted that not all of Descartes' contemporaries were satisfied with the interaction of radically different substances, either, particularly those for whom causation was understood to necessarily involve physical contact.

⁶ However, for a particularly energetic, contemporary defense of Cartesian interactionism, I recommend Foster (1991).

⁷ Robert Oates (2010) makes a strong argument for the conclusion that, with each step in its advancement, physics has become increasingly non-physical.

footprint in Western thought, including Plato (at least according to some), Leibniz, Immanuel Kant, Hume, Berkeley, Hegel, Schopenhauer and probably Edmund Husserl as well.

The perennial theme of the primacy of consciousness can also be seen in relation to an important strand in contemporary analytic philosophy that questions the traditional conceptions of the reality and externality of the physical world. Motivated by the apparent dependence of *meaning* on environmental considerations, as well as by traditional issues in skeptical epistemology, many contemporary philosophers have questioned the extent to which the physical world is objective and external.⁸ This recent development resonates with the Vedic tradition of India, where the theme of the primacy of consciousness has predominated for millennia, providing a rich and comprehensive theoretical framework together with systematic empirical methods based on the primacy of consciousness.

My own understanding has advanced in the context of this Vedic tradition (for those familiar with Indian scholarship, the view I ascribe to is known as Vedanta), and I would be remiss if I didn't acknowledge my intellectual indebtedness to the modern Vedic scholar and scientist of consciousness, Maharishi Mahesh Yogi. In addition, my understanding of how consciousness relates to the advances in modern theoretical physics has been enormously influenced by the lectures and writings of my colleague, John Hagelin.

Recent developments in quantum field theory tend toward a description of nature that integrates all the phenomena of the material world on the basis of a single, comprehensive unified field, which represents the most fundamental level of natural law. In this context, the primacy of consciousness would imply the identity of consciousness and the unified field. This idea has been advanced by Maharishi and

Hagelin since the early 1980's (Maharishi, 1986; Hagelin, 1987; 1989; 1998). From this perspective, it would not be surprising that physicalism in all its manifestations—including Jaegwon Kim's sophisticated, delicately balanced variation—has so far failed to provide a satisfactory resolution of the mind-body problem. The reason would be that neither consciousness, nor its intrinsic cognitive/intentional properties, are reducible to the localized, ultimately classical-physics brain, which is understood in terms of "bits of matter." And so, from this perspective, the fact that Kim's conception of physics as "bits of matter" is out-of-date and no longer accurate is profoundly relevant for the mind/body problem. Henry Stapp, Roger Penrose and Stuart Hameroff are among the leaders in arguing forcefully for the importance of advanced physics for the mind/body problem, and (Kelly *et al.*, 2006) provides an abundance of empirical arguments for the insufficiency of classical physicalism. Although space does not permit me to develop these ideas here, I hope to contribute to the advancement of the relevant philosophical analysis and theoretical physics in future work.

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.

⁸ I have in mind the vigorous discussions involving such leading contemporary thinkers as Hilary Putnam, Michael Dummett, Tyler Burge and John McDowell, among others. The evident entanglement between mind and "external" world, underlying these discussions, I believe may have its roots in the nonlocality which is such a central theme in advanced physics.

References

- Baars B. In the Theater of Consciousness: The Workspace of the Mind. Oxford and New York: Oxford University Press, 1997.
- Baars B. The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Neuroscience*, 2002; 6 (1): 47-52.
- Foster J. The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind. London and New York: Routledge, 1991.
- Hagelin JS. Is consciousness the unified field? A field theorist's perspective. *Modern Science and Vedic Science*, 1987; (1): 29-87.
- Hagelin JS. Restructuring Physics from its Foundation in light of Maharishi's Vedic Science. *Modern Science and Vedic Science* 1989; 3(1): 3-74.
- Hagelin JS. Manual for a Perfect Government (Chapter 3). Maharishi University of Management Press, 1998.
- Horgan T and Tienson J. The intentionality of phenomenology and the phenomenology of intentionality. In Chalmers DJ (ed). *Philosophy of Mind: Classical and Contemporary Readings*. Oxford and New York: Oxford University Press, 2002.
- Horgan T and Kriegel U. Phenomenal intentionality meets the extended mind. *The Monist* 2008; 91: 353-380.
- James W. *The Principles of Psychology*. New York: Henry Holt & Company, 1890. Available online at <http://psychclassics.yorku.ca/James/Principles/index.htm>. Accessed date: Sept 15, 2010.
- Kelly E, et al. *Irreducible Mind: Toward a Psychology for the 21st Century*. Lanham, MD: Rowman & Littlefield Publishers, 2006.
- 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.
- Maharishi Mahesh Yogi. *Life Supported by Natural Law: Discovery of the Unified Field of All the Laws of Nature and the Maharishi Technology of the Unified Field*. Age of Enlightenment Press, 1986.
- Oates, Robert. *Consciousness and the Quantum: Maharishi's Self-Interacting Dynamics of Consciousness as the Interpretation of Quantum Mechanics*. Ph.D. dissertation at Maharishi University of Management, 2010.
- Strawson G. *Mental Reality*. Cambridge MA: MIT Press, 1994.