"A defense of the explanatory argument for physicalism"

Jared Bates, Hanover College

The Philosophical Quarterly 59 (2009): 315-324.

Abstract: One argument for reductive physicalism – the explanatory argument – rests on its ability to explain the vast and growing body of acknowledged psychophysical correlations. Kim has recently leveled four major objections against the explanatory argument. Here I reconstruct and assess all of Kim's objections, showing that none is successful. The result is a defense of the explanatory argument for physicalism.

One argument for reductive physicalism rests on its ability to explain the vast and growing body of acknowledged psychophysical correlations – the set of lawlike, mental/physical covariances delivered by the neurosciences.[1] The argument can be formulated in two steps. The first is a sequence of particular applications of *inference to the best explanation*, each of which fits the following schema:

P1. $FACT_i$: Mental state-type M_i and physical state-type P_i are such that it is nomologically necessary that for all x, x is in M_i iff x is in P_i, i.e., $\forall x(M_i x \leftrightarrow P_i x)$.
P2. One possible explanation of $FACT_i$ is H_i: M_i and P_i are one and the same state-type, i.e., $M_i=P_i$.
P3. No alternative explanation of $FACT_i$ is as good as H_i.
∴ HENCE, H_i is true, i.e., $M_i=P_i$.

At this first stage, empirically confirmed psychophysical correlations are one by one offered explanations by corresponding psychophysical identities. And it is argued, by appeal to such criteria as parsimony and conservatism, that psychophysical identities are better explanations of these phenomena than rival (antiphysicalist) hypotheses. The second step in the argument is in effect an *inductive generalization* of the pattern

[1] J.J.C. Smart, "Sensations and brain processes" *Philosophical Review*, 68 (1959), pp. 141-56; Christopher S. Hill, *Sensations: a defense of type materialism* (Cambridge: Cambridge University Press, 1991); Brian P. McLaughlin, "In defense of new wave materialism: a response to Horgan and Tienson" in Carl Gillett and Barry Loewer (eds), *Physicalism and its discontents* (Cambridge: Cambridge University Press, 2001), pp.319-330; and Andrew Melnyk, *A physicalist manifesto: thoroughly modern materialism* (Cambridge: Cambridge University Press, 2003).

established in the first. It says that as the list of psychophysical identities continues to grow longer and more richly varied, then (barring counterinstances) the more probable it is that for every mental type, there is a physical type with which it is identical. Let us call the argument comprising both the steps sketched here 'the explanatory argument for physicalism,' or just 'the explanatory argument.'

Antiphysicalists have paid astonishingly little focused attention to the explanatory argument for physicalism. One exception is a series of objections against the explanatory argument recently advanced by Jaegwon Kim.[2] By my count, Kim levels four major objections against the explanatory argument. My task here is to reconstruct and assess all of these objections, showing that none is successful. The result is a defense of the explanatory argument for physicalism.

Before turning to Kim's objections, let me briefly note two points about the explanatory argument as I have formulated it, neither of which obviates Kim's objections, but both of which elaborate the argument in important ways. First, what is meant by 'physical type' in the context of this argument? While the term still has strong associations with earlier type materialist views, recent versions of the argument intend a more inclusive application. In particular, 'physical type' is meant to include functional types, all of whose tokens we can expect to be entirely physically realized.[3] One alternative way to formulate the physicalist conclusion of the explanatory argument then is: "For every mental type, there is a physical (or physically realized functional) type with which it is identical." This formulation stands neutral between type materialism and psychofunctionalism, and even allows a type materialist account of some mental types and a psychofunctionalist account of others. For our purposes here, we can take it that all of these options are versions of reductive physicalism. While this claim may be generally controversial, Kim for one would not balk at it.[4]

---

[2] Jaegwon Kim, *Physicalism, or something near enough* (Princeton, NJ: Princeton University Press, 2005). The entirety of his critique is found in pp.126-39. All subsequent page and chapter references to Kim will be to this work unless otherwise noted.

[3] McLaughlin, "In defense of new wave materialism", and Melnyk, *A physicalist manifesto*. For instance, McLaughlin uses "S=P/F" to express the identity between a "type of sensation state", on the one hand, and a "type of physical or functional state", on the other, in his formulation of the argument (p.318). The reader will have to bear in mind the inclusive sense intended by my use of 'physical type'.

[4] Kim, ch.4, and Jaegwon Kim, "The mind-body problem: taking stock after forty years" *Philosophical Perspectives*, 11 (1997), pp.185-207. For a slightly different account of why psychofunctionalism entails a robust form of reductive physicalism, see Melnyk, *A physicalist manifesto*, ch.3.

Second, the claim that psychophysical identities offer better explanations of psychophysical correlations than rival hypotheses is supported in three ways, and these need at least to be sketched. First, none of the antiphysicalist theories *can* explain particular psychophysical correlations. On any sort of dualism, psychophysical correlations must be accepted as brute, lawful associations between ontologically diverse types, and so the explanatory advantage tips initially to physicalism for offering any explanation at all.[5] Moreover, physicalist explanations postulate fewer entities and properties than any dualist explanation does, and so physicalist explanations are preferable on grounds of parsimony or economy.[6] Finally, physicalist explanations cohere with antecedent commitments to other beliefs, like belief in the causal closure of the physical, the causal efficacy of the mental, and the belief that actions are not typically overdetermined by distinct causal pathways. Kim himself acknowledges that physicalism solves (in a certain sense) the problem of mental causation,[7] though he does not ultimately endorse global physicalism. Given this, physicalist explanations can claim a measure of conservatism, or a better fit with background commitments, over antiphysicalist rivals.[8] With the balance of explanatory strengths on their side, physicalist explanations are preferable to alternative explanations.

While some of the claimed explanatory strengths of physicalism will strike many as contentious – Kim himself directly attacks the first one – we will await the occasion to revisit them below. For now we turn to consider Kim's four objections to this argument. The first three of these call into question inference to the best explanation both generally and in its specific application to the mind-body problem. The final, and strongest, objection argues that psychophysical identities do not explain psychophysical correlations at all, potentially placing physicalism and dualism on an equal explanatory footing. Below I reconstruct and assess each of these objections, concluding that the explanatory argument can be defended at every turn. The unfolding discussion shows that the explanatory argument presents a formidable argument for reductive physicalism.

---

[5] Hill, *Sensations*, pp.24-6, and Melnyk, *A physicalist manifesto*, p.301.

[6] Hill, *Sensations*, pp.39-40; Hill and McLaughlin, "There are fewer things in reality than are dreamt of in Chalmers's philosophy" *Philosophy and phenomenological research*, LIX (1999), pp.445-454 (esp. 449 and 451); McLaughlin "In defense of new wave materialism", p.320; and Melnyk, *A physicalist manifesto*, pp.244ff.

[7] Kim, p.55.

[8] Hill and McLaughlin "There are fewer things in reality…", p.451, and McLaughlin, "In defense of new wave materialism", p.320.

> Objection 1: Even if it is granted that reductive physicalism gives better explanations of the phenomena than nonreductivist rival hypotheses, the inference to the best explanation is nevertheless insufficient to warrant our acceptance of reductive physicalism as true. Hence, we are under no rational compulsion to accept the conclusion of the explanatory argument.[9]

This objection questions the specific connection between physicalism's explanatory superiority over rivals, on the one hand, and our warranted acceptance of it, on the other. One reason for thinking the connection so weak in the present case is that there is generally a weak connection between a theory's explanatory credentials and its truth or probability. Although Kim does not *explicitly* express such a general skepticism, he does point the reader to van Fraassen's criticisms of inference to the best explanation, which notoriously do express such skepticism.[10] Perhaps Kim is presupposing something like the following principle:

> (EP) If a hypothesis h's explanatory superiority over rivals does not entail that h is at least probable (p(h)>.5), then h's explanatory superiority over rivals is insufficient to warrant our acceptance of it.

But (EP) is a dubious epistemic principle. Suppose our epistemic end is to accept as comprehensive and error-free a picture of the world as possible. Against this end, we can evaluate different possible cognitive attitudes we might take on some matter of explanatory interest. For instance, *withholding all judgment* on some matter is a perfect safeguard against error, but it also advances not a step toward comprehensiveness. Alternatively, *disbelieving available explanations* minimizes the risk of error (though not entirely) if the available explanations all have low probabilities (since $p(\neg h)>.5$ if $p(h)<.5$). This attitude comports with (EP), but again fails to move much toward comprehensiveness. By contrast, *believing the most likely explanation*, no matter how low its probability, maximizes comprehensiveness while minimizing the risk of error as far as it can. Viewed this way, rationally accepting an explanation does not require assigning it a high probability.[11] The received knowledge of current science is probably

---

[9] Kim, pp.128-9. Note that I reconstruct Kim's objections in displayed paragraphs, and that these are not direct quotes.

[10] Bas van Fraassen, *Laws and symmetry* (Oxford: Clarendon Press, 1989), cited by Kim on p.131,n.22.

[11] Mark Kaplan, *Decision theory as philosophy* (Cambridge: Cambridge University Press, 1996), ch.4.

best understood as taking this third option, for this knowledge largely comprises hypotheses whose explanatory credentials fall short of making them more probable than not; yet it is precisely their explanatory credentials that warrant our acceptance of them. What this means is that if physicalism's explanatory superiority over its rivals is granted (as this objection does grant), then accepting physicalism is epistemically rational.

> Objection 2: Inference to the best explanation, while appropriate in the empirical sciences, is out of bounds in disputes between philosophical theories in general and between competing theories on the mind-body problem in particular. Hence, we are under no rational compulsion to accept the conclusion of the explanatory argument for physicalism.[12]

In support of this view, Kim points to the ongoing nature of scientific testing with its tentative, revisable results and contrasts this with the prima facie implausibility of settling philosophical disputes by inductive methods. While much could be said about how this objection characterizes the actual practice of science and philosophy, this line of thought can be subverted by the reply to the last objection: We are warranted in taking the best means to our epistemic ends. So if in some context inference to the best explanation is the most suitable method for reaching those ends, we are permitted to use it. In light of this, harboring a soft spot for the autonomy of philosophy, with its insulation from empirical discoveries and rejection of inductive modes of reasoning, unduly cripples our intellectual pursuits. Why *might* inference to the best explanation be suitable for reasoning about minds and bodies? Psychophysical correlations *do* have a bearing on how we think about mentality and how it fits with the rest of the world. (To see this, just imagine the plausibility of physicalism had investigation revealed the brain to be the undifferentiated gray blob it initially appears to be.) Moreover, since psychophysical correlations are empirical discoveries, inference to the best explanation may play *some* role in our drawing conclusions from them, just as it does with other empirical discoveries. If there is a mistake in how the explanatory argument for physicalism draws out its conclusions, then that mistake needs to be pointed out. The explanatory argument cannot be rejected just because it is an explanatory argument.

> Objection 3: As an inductive method, inference to the best explanation is beholden to the rule of total evidence, that is, it must take into account *all* of the relevant data. By arbitrarily focusing only on the psychophysical

[12] Kim, p.130.

> correlations, the explanatory argument runs afoul of the rule of total evidence. Hence, we are under no rational compulsion to accept the conclusion of the explanatory argument.[13]

What other data should be brought to bear in an inference to the best explanation for a theory of mind? Kim asserts that a theory should give the best explanation of "all the facts in the domain"[14] and these include not only the psychophysical correlations, but also "the presumptive authoritativeness and privacy of first-person access to one's own mental states, the persistence conditions of persons, the multiple physical realizability of mental properties, the possibility of qualia inversions, the possibility of 'zombies,' and the like".[15] But Kim's application of the rule of total evidence is mistaken for two independent reasons. First, the rule of total evidence cannot possibly require that hypotheses literally give explanations of all the facts within a domain. For instance, hypotheses about the thermal properties of ascorbic acid do not have to explain all the facts about ascorbic acid or all the facts about thermal chemistry. That said, while hypotheses about the thermal properties of ascorbic acid do not have to explain (because they cannot explain) all the facts about ascorbic acid and thermal chemistry, perhaps it is a mark in their favor that they cohere with other things we already know about ascorbic acid and thermal chemistry. In other words, we should understand the *rule of total evidence* in inference to the best explanation as nothing more than an appeal to *conservatism* as an explanatory criterion: Other things being equal, the more conservative explanation is to be preferred over less conservative ones. As discussed earlier physicalists do in fact argue for identifying mental types with physical types on grounds of conservatism. One specific such claim is that reductive physicalism gives the most conservative solution to the problem of mental causation – conservative because it (and it alone) relaxes the tension between our commitments to the causal efficacy of the mental, the causal closure of the physical, and the resistance to rampant causal overdetermination. This line of thought is sometimes known as the 'causal argument' for physicalism,[16] though Kim presents it as the problematical 'supervenience argument' against nonreductive physicalism.[17]

---

[13] Kim, p.129.

[14] Kim, p.129,n.20.

[15] Kim, p.129.

[16] For two recent presentations, see David Papineau, *Thinking about consciousness* (Oxford: Oxford University Press, 2002), ch.1; and Andrew Melnyk, *A physicalist manifesto*, pp.285-97, reprinted in slightly modified form as "Some evidence for

Before turning to the second problem with Kim's application of the rule of total evidence, we have to deal with the potential difficulty that the appeal to mental causation renders the explanatory argument superfluous.[18] The causal argument alone is supposed to *entail* physicalism. But then it seems mental causation is doing all the work for the explanatory argument. Why then is the explanatory argument needed? The answer is that the explanatory argument and the causal argument present two species of empirical evidence for physicalism linked arm-in-arm together. That the causal argument is sometimes presented as a deductive argument can obscure this fact, though I would hasten to add that a deductive argument does not trump an inductive argument just because it is deductive. David Papineau presents the causal argument as deductive. Summarizing:

(1) Mental states have physical effects.
(2) All physical effects are fully caused by purely physical prior states.
(3) The physical effects of mental states are not overdetermined by distinct causes.
(4) Hence, the mental states mentioned in (1) are identical with some part of the physical causes mentioned in (2).[19]

However Papineau supports all three of these premises by appeal to empirical evidence and standard explanatory criteria. In particular, (3) is supported by the fact that science has found no analogous cases of overdetermination elsewhere in nature.[20] Compare this with Andrew Melnyk's presentation of the causal argument, in which (4) is offered as the best explanation of (1) and (2). The non-overdeterminationist premise (3) is supplanted by appeals to analogy and economy on (4)'s behalf.[21] As I see it, these two versions of the argument, one deductive and one not, confer equal justification on

---

physicalism" in *Physicalism and mental causation*, ed. by Sven Walter and Heinz-Dieter Heckmann (Charlottesville, VA: Imprint Academic, 2004).

[17] Kim, ch.2. The general argument is advanced by Kim in a number of places, at least as far back as "The myth of nonreductive materialism" *Proceedings and addresses of the American Philosophical Association*, 63 (1989), pp.31-47 (reprinted in his *Supervenience and mind* (Cambridge: Cambridge University Press, 1993), ch.14). See also his *Mind in a physical world* (Cambridge, MA: MIT Press, 1998), ch.2.

[18] This objection is due to an anonymous referee for this journal.

[19] Papineau, *Thinking about consciousness*, pp.17-18.

[20] Papineau, Thinking about consciousness, pp.27-28.

[21] Melnyk, "Some evidence for physicalism", pp.158-162.

reductive physicalism. Moreover, the degree of justification the argument confers on physicalism is not so great as to render the explanatory argument otiose. Rather, mental causation and psychophysical correlations, though interlinked, both provide empirical support for physicalism.

The second problem with Kim's application of the rule of total evidence concerns his charge that focusing on the psychophysical correlations to the exclusion of other evidence is "arbitrary". Two points undermine this charge. First, as Kim himself says of the above-quoted list of things a theory of mind must explain, "[t]hese are all contested and disputed issues".[22] Indeed, claims of the possibility of zombies and inverted qualia are among the most disputed claims in recent philosophy of mind.[23] Second, and by contrast, the psychophysical correlations are generally *not* in dispute. According to Kim, all parties to the mind-body debate – epiphenomenalists, emergentists, double-aspect theorists, etc. – agree that mental types at least nomologically covary with physical types.[24] Here we see why focusing on psychophysical correlations in an inference to the best explanation is not at all arbitrary: They are some of the only things in this domain that are not in dispute. An explanation is only needed when some acknowledged phenomenon occurs, and the psychophysical correlations are acknowledged by all parties. An explanatory argument may take any acknowledged phenomenon as its explanandum, and more generally any theory may take as its starting-point the common ground it shares with competing theories. Appeal to the rule of total evidence, as far as I can make it out, does not change the overall shape or strength of the explanatory argument for physicalism.

Objection 4: Psychophysical identity-statements do not explain psychophysical correlations. So the second premises in all the iterations of inference to the

---

[22] Kim, p.129. Though it has to be added that Kim grants physicalism an account of most of the items on his preferred list. He is explicit that multiple realizability poses no problem for reductive physicalism, p.164. He also argues that intentional psychological states (beliefs, desires, etc.) are functionally reducible to the physical, pp.165-8. And in his textbook, *Philosophy of mind*, 2nd ed. (Cambridge, MA: Westview Press, 2006), pp.211-6, he gives a quasi-functionalist account of privileged access as part of our intentional psychology, brushing it aside as a defining feature of phenomenal consciousness. Note also that mental causation is conspicuously missing from the list.

[23] There is a vast literature here, but to limit myself to sources already cited see Hill and McLaughlin, "There are fewer things in reality …", Melnyk, *A physicalist manifesto*, ch.1, and Papineau, *Thinking about consciousness*, ch.3.

[24] Kim, p.34.

> best explanation in the explanatory step of the argument are false. Hence, we are under no rational compulsion to accept the conclusion of the explanatory argument.[25]

This is the objection that exercises Kim the most. We will look at two different attempts he makes at showing that psychophysical identity-statements do not explain psychophysical correlations.

> First attempt: If psychophysical identities explain psychophysical correlations, then it would be possible to *derive* a correlation-statement directly from its corresponding identity-statement. But this is not possible.[26]

Recall that, as I have formulated the schematic inference to the best explanation in the explanatory step of the argument, the explanandum is that, for some M and some P, $\forall x(Mx \leftrightarrow Px)$, and the physicalist explanans is that M=P. Kim's point is that it is impossible to derive '$\forall x(Mx \leftrightarrow Px)$' from 'M=P' alone, for as he says, "To my knowledge, there is no rule in logic that says 'From 'X=Y,' infer 'X occurs $\leftrightarrow$ Y occurs''".[27] The only way possible to run such a derivation, according to Kim, is to assume both the psychophysical identity-statement, M=P, and the tautologous biconditional, $\forall x(Mx \leftrightarrow Mx)$. Then the psychophysical correlation will follow by the substitutivity of identity:

1.	M=P	A
2.	$\forall x(Mx \leftrightarrow Mx)$	A
3.	$\forall x(Mx \leftrightarrow Px)$	1,2 substitutivity rule

This is surely no explanation, says Kim, for the third line of the derivation is "simply a rewrite" of the "factually empty logical truth" in the second line.[28] It would, of course, be a mistake to think that derivations *suffice* for explanations. The point Kim insists on is that, assuming Hempel's model, derivations are *required* for explanations. However, Kim is mistaken about two things: The first is that the derivation required for an explanation should in any way be illuminating. Granting near-completeness to a system for second-

---

[25] Kim, pp.132-9.

[26] Kim, pp.134-6.

[27] Kim, p.135.

[28] Kim, p.135. Kim characterizes the contributions of identity-statements to explanations as *rewriting* or *redescribing* facts multiple times (once on p.133 and p.134, and twice on p.132 and p.135).

order validity, it is a corollary of Hempel's model (granted for argument's sake) that if K explains S, then there is a derivation of S from K. But such a derivation is merely a set of syntactic manipulations on the explanans. While Kim is right that the substitutivity rule does not assist in explanation, it has to be pointed out that neither do any of the other rules. In this sense, *modus ponens* is as much a "rewrite rule" as the substitutivity rule – what gets rewritten is the consequent of a given conditional. On Hempel's model, the crucial feature of an explanation is the entailment-relation between the explanans and explanandum; the derivability-relation is a corollary, not a constituent. Secondly, and more importantly, while Kim may be right that there is no derivational system in print with a single rule allowing you to turn identity-statements into biconditionals, such a rule would certainly be valid. Adding such a rule to any of the standard derivational systems would then neutralize Kim's objection. But adding the rule is unnecessary, for there are already ways to derive the wanted biconditional from the identity-statement alone:

1.	M=P	A
2.	Ma	A
3.	Pa	1,2 substitutivity rule
4.	Pa	A
5.	Ma	1,4 substitutivity rule
6.	$Ma \leftrightarrow Pa$	2-3,4-5 biconditional proof
7.	$\forall x(Mx \leftrightarrow Px)$	6 universal generalization

Kim's objection presupposes that the derivability of the explanandum from the explanans is a necessary condition on explanation. As shown above, line 7 is derivable from line 1 as its sole primary assumption, and is not merely a rewrite of an additional, explanatorily empty tautology. So Kim is mistaken about the derivability of biconditionals from identity-statements, and this objection does not lay any demands on explanation that physicalism cannot meet.

Second attempt: A genuine explanation moves us from one fact to a different fact. But identity-statements merely allow us to redescribe facts we already

know under other descriptions. So identity-statements have no explanatory power of their own.[29]

As noted above, Kim relegates identity-statements to the role of "rewrite" rules, with the implication that rewriting does not make for explanation. He illustrates this point with the following argument:[30]

[1] Tully is wise.
Tully = Cicero.
Therefore, Cicero is wise.

The argument here moves from a fact under one description (*Tully's being wise*) to the same fact under a different description (*Cicero's being wise*). As such the argument fails as an explanation of *why* Cicero is wise. Moreover, Kim counts as his allies in advancing this objection such reductive physicalists as Smart, Feigl, and Block and Stalnaker,[31] all of whom agree that psychophysical identities supplant psychophysical correlations and thereby stamp out any need for explanation.

The first thing to say here is that Kim's physicalist cohorts do not at all argue against the notion that psychophysical identities explain psychophysical correlations, which is what Kim is attempting to show. Rather, all are arguing against some or other form of dualism, and each explicitly argues for the preferability of psychophysical identities over *irreducible correlation-laws*, which is really what gets supplanted by the identities. All of this is of course in tune with the explanatory argument.[32]

---

[29] Kim, pp.132-4 and pp.136-9.

[30] Kim, p.132.

[31] Smart, "Sensations and brain processes", Herbert Feigl, "The 'mental' and the 'physical'" *Minnesota Studies in the Philosophy of Science*, 2 (1958), pp.370-497, and Ned Block and Robert Stalnaker, "Conceptual analysis, dualism, and the explanatory gap" *The Philosophical Review*, 108 (1999), pp.1-46.

[32] Once the psychophysical identities are vouchsafed, of course, the advocate of the explanatory argument can agree with Block and Stalnaker that no *further* explanations of the phenomena are needed – especially not the sort of a priori conceptual explanations with which they are concerned. These are the "reductive" explanations demanded by David Chalmers, *The conscious mind* (New York: Oxford University Press, 1996), Frank Jackson, "Finding the mental in the natural world" in R. Casati, B. Smith, and G. White (eds), *Philosophy and the cognitive sciences* (Vienna: Verlag Hölder-Pichler-Tempsky, 1994), and Joseph Levine, "On leaving out what it is like" in M. Davies

But let us return to Kim's 'Tully = Cicero' argument, [1]. Even if it is granted that [1] above does not constitute an explanation, Kim generalizes too hastily to the conclusion that identity-statements are never explanatory. To see this, compare [1] with

[2] Tully = Cicero.
Therefore, Tully is around iff Cicero is around.

This argument clearly moves from one fact (*Tully's being identical with Cicero*) to a different fact (*the constant joint appearance of Tully and Cicero*), and so, once again, it satisfies Kim's requirements on explanation. But more has to be said here, for the argument goes well beyond satisfying Kim's requirement that it move from one fact to another. That is, the argument surely gives the *correct* explanation of Tully's and Cicero's going around together. Simply put, once we understand that Tully *just is* Cicero, it is no wonder that wherever Tully goes, there goes Cicero. In fact, no correct explanation of this fact could fail to mention that Tully and Cicero are one and the same.

The explanation in [2] is directly relevant to the case for physicalism, for we just showed that [2] gives an adequate explanation of a correlation in terms of an identity. If Tully's being identical with Cicero explains their correlated appearances, then certainly psychophysical identities equally explain psychophysical correlations. That is, if we acknowledge that a mental type, M, and a physical type, P, are one and the same, then the lawful covariance of M and P is no surprise, i.e., we expect that people are in M iff they are in P. The conclusion to draw here is that, once again, Kim has not placed any demands on explanation that physicalism cannot meet. Just because identity-statements are not always explanatory (as in [1]) does not mean that they never are. As we have seen in [2], identity-statements are sometimes indispensable in giving correct explanations.

Neither of Kim's attempts to show that psychophysical identities do not explain corresponding psychophysical correlations is successful. The only necessary conditions Kim lays on explanation – namely, that the explanandum be derivable from the explanans and that the explanation move from one fact to another, rather than to a mere rewrite of the first fact – are conditions physicalist explanations more than satisfy. The second premise in the schematic inference to the best explanation remains entirely secure. That premise makes the minimal assertion that one possible explanation for $\forall x(Mx \leftrightarrow Px)$ is that M=P. Kim has not given a reason to doubt this.

---

and G.W. Humphreys (eds), *Consciousness: psychological and philosophical essays* (Oxford: Blackwell Publishers, 1993), all of apiece with Kim's own model, ch.4.

In fact, none of Kim's objections against the explanatory argument has been successful. This is surely a sign of strength for the explanatory argument, though it is not of course a conclusive proof of reductive physicalism. The evidence accrued to physicalism by the explanatory argument is both incremental and defeasible. It is incremental insofar as the evidence gets stronger as more psychophysical identities are confirmed (barring genuine counterinstances). And it is defeasible insofar as the evidence is nevertheless logically compatible with physicalism's falsity. But the argument is formidable. Antiphysicalists who resist its conclusion owe an account of where they think the argument goes wrong. There has been little critical discussion of the explanatory argument to date. Kim's attacks take aim at (i) inference to the best explanation itself (by arguing that this mode of reasoning cannot warrant our acceptance of physicalism as a correct theory), (ii) the legitimacy of appealing to inference to the best explanation in the present context (by arguing that inductive reasoning is outside its jurisdiction in philosophical debates), (iii) the legitimacy of focusing exclusively on the psychophysical correlations (by arguing that this runs afoul of the rule of total evidence), and (iv) the second premises in the applications of inference to the best explanation (by arguing that psychophysical identities do not explain psychophysical correlations). I have shown that the explanatory argument for physicalism can be defended at every turn. The result is that the explanatory argument gives us good, empirical grounds to accept physicalism. Antiphysicalists are warmly invited to accept this result.[33]

[33] I would like to thank Jim McBain, Andrew Melnyk, Jerry Vision, and two anonymous referees for this journal for helpful comments on an earlier draft of this paper.