Jaegwon Kim, *Physicalism, or something near enough*

Princeton, NJ : Princeton University Press, 2005,
200 pp., $26.95, ISBN : 0-691-11375-0.

(Review/Recension/Rezension)

Christian Sachse

Université de Lausanne
Centre romand de logique, histoire, et philosophie des sciences

christian.sachse@unil.ch

publié à / published by / veröffentlich in :
Dialectica, 60: pp. 508-512.

Le travail nécessaire à la réalisation de cet article a été soutenu par le Fonds National de Suisse (FNS), bourse no. 100011-105218/1
The work on this paper has been supported by the Swiss National Science Foundation (SNF), grant nr. 100011-105218/1
Die Arbeit an diesem Artikel ist unterstützt durch den Schweizer Nationalfonds (SNF), Stipendiums-Nr. 100011-105218/1

Jaegwon Kim
Physicalism, or something near enough

Review by
Christian Sachse
Université de Lausanne

In his new book, Kim gives an overview of his influential ideas on the mind-body debate. Within this debate, a non-reductive physicalism has been the standard position during the last decades. However, due to Kim's work, something of a revival of reductive physicalism has gradually taken place. Nowadays, it seems to be the case that many philosophers recognize that mental properties are either deducible *or* epiphenomenal. In general, Kim sets out to assess how much of physicalism is true. What are the consequences and what are the limits of his famous supervenience argument? First, I give an overview of the topics Kim deals with. I then comment on identities, explanation and his final conclusion – 'Physicalism is not the whole truth, but it is the truth near enough, and near enough should be good enough' (174).

Kim starts by sketching his position within the mind-body debate. He emphasises that mental causation and consciousness are indispensable for our self-conception. Human agency and human knowledge presuppose that our mental states are causally efficacious. 'If a phenomenon is to have an explanatory role, its presence or absence must make a difference – a *causal* difference' (10; italics in the original). But this is a challenge to contemporary physicalism: How can minds exert their causal powers in a world that is fundamentally physical and causally closed? How can we give a physicalist account of consciousness? To understand the problem of mental causation, let's turn to chapter 2. There, we can take up the standard position of non-reductive physicalism that can be summed up as follows:

1. 'Mental properties strongly supervene on physical/biological properties. That is, if any system *s* instantiates a mental property M at *t*, there necessarily exists a physical property P such that *s* instantiates P at *t*, and necessarily anything instantiating P at any

time instantiates M at that time' (33).

2. 'Mental properties are not reducible to, and are not identical with, physical properties' (34).
3. 'Mental properties have causal efficacy – that is, their instantiations can and do cause other properties, both mental and physical, to be instantiated' (35).

Although there are good reasons for each principle, together they are incompatible with the following two principles:

4. Closure of the physical domain: 'If a physical event has a cause that occurs at *t*, it has a physical cause that occurs at *t*' (43).
5. Causal exclusion: 'No single event can have more than one sufficient cause occurring at any given time – unless it is a genuine case of causal overdetermination' (42).

Thus, the causal efficacy of mental properties seems to be pre-empted by physical causes. So, Kim's supervenience argument points out the inconsistency of both mental causation *and* mental irreducibility. The only possibility to save mental causation seems to be a reductive position.

Following the reductionist consequences of the supervenience argument, Kim criticises the dualist approaches in chapter 3. He argues against a revival of substance dualism by setting out in detail its, above all *causal*, problems in our physical world. Any cause requires a spatiotemporal location in order to be distinguishable from other causes. However, is there a clear idea of how 'a wholly nonspatial mental structure could account for a soul's causal power' (90)? Every causally efficacious property has to be physically located (and therefore physical).

In chapter 4, Kim deals with the closely connected topics of Nagel's bridge-law reduction, his favoured model of functional reduction and Levine's explanatory gap. The problem is that Nagel's bridge-law reduction lacks explanatory power. In addition to this failure, Nagel's bridge-law reduction faces the multiple-realizations problem: Mental properties are often (if not always) multiply realized. This means that a mental property M is realized by different physical properties (say P_1 and P_2). Therefore, M can neither be bi-conditionally connected with P_1 nor with P_2. Hence, no reduction of M seems to be possible by means of biconditional

bridge-laws. However, consider the following three steps of Kim's preferred functional model

of reduction:

I. Functionalization of the mental property: 'Property M to be reduced is given a *functional definition* of the following form:

Having M = def. having some property or other P (in the reduction base domain) such that P performs causal task C' (101, italics in the original).

II. Discovery of the realizers of M: 'Find the properties (or mechanisms) in the reduction base that perform the causal task C' (102).

III. Explanation of M: 'Construct a theory that explains how the realizers of M perform task C' (102).

Following this model, M can be functionally reduced to P and reductively explained. Therefore, the functional model of reduction enables an explanatory ascent from the basis level of P to the functional level of M: The realizer theory explains how P satisfies the causal task C that functionally defines M. Thus, M is reductively explained in terms of the realizer theory of P.

Linking up with the previous chapter, Kim elaborates in chapter 5 on positive arguments for type identities. According to Hill and McLaughlin, type identities offer the best explanation of the mind-body correlations. But, according to Kim, identities 'seem best taken as mere rewrite rules in inferential contexts' (132) and for that reason they 'seem not to have explanatory efficacy of their own' (132). He agrees with Block and Stalnaker that any correlation upgraded to an identity leaves nothing to be explained. Nevertheless, to be causally efficacious in a physical world that is causally closed, mentality must be part of it – the mental states must be identical with brain states. And, to explain the causal efficacy of M is to explain the lower-level mechanisms that fulfil the causal task of M (at least at the level of instances supposing multiple realization). Thus, it is a *causal* argument that provides reasons for moving from correlations to identities – and not a purely *explanatory* argument.

In chapter 6, Kim takes stock and discusses the results of the previous chapters – and doubts whether or not mentality is reducible and can be captured by a physicalist worldview. For him, there are mental properties that can be captured by a functional definition – but on the other hand, there are the qualitative aspects of some mental properties that resist functionalizability and therefore remain irreducible. Thus, all the reducible mental properties belong to the physical domain and are causally efficacious. This saves cognition and agency, and for Kim it is truth near enough and good enough. However, the qualitative aspects of

qualia are not functionalizable and, hence, irreducible. As a result, they remain epiphenomenal and outside of the physical domain. That's why Kim takes physicalism not to be the whole truth.

Here is a difficulty for Kim's reductionism: he argues that type identities have no explanatory role of their own. Now, let us suppose a multiple realization of a higher-level property type M. Consequently, there are different realizer types (say P_1 and P_2) that are causally heterogeneous. 'If the term "multiple" in "multiple realizations" means anything, it must mean causal/nomological multiplicity; if two realizers of pain are not causally or nomologically diverse, there is no reason to count them as two, not one' (26). However, if we have causally heterogeneous realizers, how is a homogeneous reductive explanation of the higher-level property type possible? Whether P_1 or P_2 is the realizer of M makes a difference in the reductive explanation. For that reason, it seems to be necessary to establish some kind of type identity if we want to give type explanations in a homogeneous way. If only higher-level tokens were reductively explainable, this would lead to eliminative consequences about higher-level types. Then type identities might not be explanatory on their own – but they are nevertheless necessary for a homogeneous explanation of higher-level property types and thus a conservative reduction.

Here is a proposal: according to Kim, 'multiple realization only leads to reducibility to multiple reduction bases, not to irreducibility' (56). In his *Mind in a Physical World* he argued in more detail that there is a disjunction of realizer types to which we reduce the multiply realized higher-level type.[1] However, what can be reduced to a disjunction of base property types and what does this mean? In my opinion, multiple realization essentially shows that there is something imprecise at the higher-level. I claim that the causally diverse realizer types (P_1 and P_2) are in theory always distinguishable on the higher-level (of M). There is always an environment possible in which P_1 and P_2 differ in some respects that are mentally relevant. Imagine a pain (M) that takes a little bit longer to react to medicine if it is realized by P_1 and not by P_2. These functional differences of M occur in certain environments if M is realized by *different* P's. As a result of this, the different effects of the different realizer types can be taken into account in a more precise definition of the mental property. We can introduce functional subtypes of M (M_1 and M_2) within the higher-level, functional theory to establish some kind of type identity: $M_1 = P_1$, and $M_2 = P_2$. This allows for homogenous reductive explanations of the sub-types (M_1 and M_2). M is, if multiply realized, something like

[1] Jaegwon Kim 1998, *Mind in a Physical World*, Cambridge, MA: MIT Press/Bradford Books, pp. 106–112.

a practical abstraction from more detailed sub-types. Multiple realization makes it obvious that in everyday contexts we typically deal with a certain descriptive imprecision. Nonetheless, it is possible to reach conservative reduction for a multiply realized property type by means of its functional subtypes: the sub-types are identical with their realizer types and therefore reducible. Moreover, the difference between M and its sub-types can be formulated within the theory of M. Each of these sub-types reveals the homogeneous structure that characterizes M.[2]

A final remark: Whether or not inverted qualia, absent qualia, or zombies are possible, Kim's reductive implications of the supervenience argument lead to the following consequence: if some aspects of our mentality are not functionalizable, they remain irreducible but causally impotent. Thus, they cannot have a place in causal explanations. Against a dualist view he states that 'immaterial nonspatial minds would be totally causally impotent, and this renders them explanatory irrelevant and useless. Moreover, such a radically noncausal view of minds makes it difficult to understand how we could even come to know that there are minds'
(151).

But what about the irreducible aspects of qualia? How can we take qualia seriously if they have no causal and explanatory relevance? That a king is aware of a problem within his realm makes necessary further commitments. Why not be more optimistic about the functionalizability of qualia? For instance, it might turn out to be possible to introduce something like functional sub-types of qualia. These sub-types of qualia enable us to give reductive explanations that take into account the possible differences between states of qualia. Then, our *whole* mentality has a real place in our physical world, for we can functionally reduce *all* its aspects. In my opinion, Kim is only near to the truth – but not near enough.

[2] Compare this reductionist strategy with Christian Sachse 2005, 'Reduction of Biological Properties by means of Functional Sub-Types', *History and Philosophy of the Life Sciences,* **27**: pp. 427–441.