Philosophical Perspectives 11 (1997): 133-48.

Kim’s Functionalism

Marian David
Notre Dame University

In some recent articles, Jaegwon Kim has argued that non-reductive physicalism is a myth: when it comes to the mind-body problem, the only serious options are reductionism, eliminativism, and dualism.[1] And when it comes to reductionism, Kim is inclined to regard a functionalist theory of the mind as the best available option—mostly because it offers the best explanation of mind-body supervenience. In this paper, I will discuss Kim’s views about functionalism. They may be contended on two general grounds. First, some functionalists will object to being classified as reductionists. Second, Kim argues for a version (or a reading) of functionalism, conceptualized functionalism, that makes it rather similar to the “old” mind-body identity theory it was designed to replace. Moreover, Kim’s conceptualized functionalism turns out to be a somewhat surprising brand of reductionism—a reductionism with some eliminativist cut-outs and, possibly, some dualist leftovers. At the end of the paper I propose a construal of the more standard version of functionalism that obviates Kim’s argument for switching-over to his conceptualized version.

1. Hard-Core Functionalism

I will sketch a form of functionalism, “hard-core functionalism”, to set up a contrast to “Kim functionalism”. Hard-core functionalism is a theory of mental states in the sense of state types or properties. (For present purposes, I will not distinguish between state types and properties.) According to hard-core functionalism, mental states (properties)

[1] See Kim 1989, 1992, and 1993.

are functional states. That means that mental states are certain second-order properties specified in terms of the *functional roles* of certain first-order properties. To use the standard example, for an organism to be in pain is for it to be in some internal state that is typically caused by tissue damage and that typically causes groans and winces and other characteristic behavior. The internal state in question here is a first-order state ϕ that plays the causal role of an internal mediator between sensory input and behavioral output. The causal role of ϕ is given by a set of input states each of which is (typically) causally sufficient for going into state ϕ together with a set of output states such that being in ϕ is (typically) causally sufficient for going into each output state. Still simplifying, the resulting causal pattern can be represented in the form of a condition $\mathrm{I}\rightarrow\phi\rightarrow\mathrm{O}$ in which 'ϕ' is a variable ranging over first-order states. This condition abbreviates the functional role of ϕ.[2]

All functionalists, of course, hold that functional roles are in some sense crucial to mental states. However, part of what makes the hard-core functionalist hard-core is that he *identifies* mental states with existential generalizations over first-order states that play the relevant functional roles. Consider, for example, a biconditional of the form:

(C) x is in pain $\leftrightarrow (\exists\phi)(x \text{ has } \phi \,\&\, \mathrm{I}_{[\text{damage}]}\rightarrow\phi\rightarrow\mathrm{O}_{[\text{wincing}]})$.

This biconditional expresses a lawful correlation between an organism's being in pain and it's having some property that is caused by tissue damage and causes wincing. Assuming the simplifications can be repaired, the hard-core functionalist considers

[2] An internal state that mediates between external input and external output may do some of its job via other internal states. This means that the specification of the functional role associated with a single mental state is likely to be vastly more complex than I have indicated. Indeed, it seems rather likely that functional roles are so heavily intertwined that a complete functional definition of one mental-state type must yield functional definitions of all mental-state types. But there is no need to go into these complications here.

such a correlation biconditional to be sufficiently strong to warrant its "upgrading" to a theoretical identification. On this view, (C) underwrites the identification of the property of being in pain with the second-order property expressed on the right hand side of the biconditional.[3] In general:

(HCF) The property M = the property of having some (first-order) property ϕ, such that being in I causes one to have ϕ and having ϕ causes one to go into O.

Mental state M is identified here with a *functional state*. That is, M is identified with a second-order property that is specified in terms of a functional role, namely the functional role "characteristic of M", as one might say. Of course, this latter reference to M is merely a crutch to facilitate talking. In the official account, the functional role "characteristic of M" is spelled out in terms of input/output conditions that make no reference to M.

So far, nothing has been said about physicalism. In fact, a functionalist does not have to commit himself to physicalism at all. However, the hard-core functionalist does want to commit himself to physicalism. To do so, he adds a further requirement, or rather, a further thesis: All mental states are *realized* in physical states. This notion of "realization" can be spelled out in the following manner. Consider the functional role $I \rightarrow \phi \rightarrow O$ that enters into the definition of mental state M: a first-order state ϕ realizes M in an organism S just in case S has ϕ and ϕ occupies the functional role (i.e., satisfies the condition $I \rightarrow \phi \rightarrow O$). So the hard-core functionalist is a physicalist because he identifies mental states with functional states that are realized in first-order physical states. This shows why, and how, mental states supervene on physical states.

[3] Second-order properties should not be confused with second-level properties. Second-level properties are properties of properties. Second-order properties are properties of individuals just like first-order properties. But unlike first-order properties, second-order properties quantify over first-order properties. Consider the property of being an Austrian and compare it with the property of *having* a quaint property (second-order) and the property of *being* a quaint property (second-level).

Moreover, since a functional role can be occupied by more than one physical state, the account seems to allow that a mental state can have multiple realizations. This is a welcome consequence. After all, it was the possibility of multiple realizations of mental states that caused the demise of the old identity theory and the rise of functionalism.[4]

The old Feigl-Smart identity theory was a type-type identity theory; it identified mental states with first-order physical states, thereby foreclosing the possibility of multiple realizations. Hard-core functionalism is also type-type identity theory. Even though functionalists are sometimes uneasy with the label 'identity theory', hard-core functionalists can hardly object to it, for they identify mental states with functional states. But we should keep firmly in mind that the two identity theories identify mental states with different kinds of states. Roughly, the old identity theory identified a mental state with a physical state which, according to hard-core functionalism, is only one of the mental state's possible realizers.

Is hard-core functionalism a form of reductionism? Yes, it reduces mental states to functional states. But is it a form of physicalist reductionism? Obviously, this depends on the nature of functional states, more specifically, it depends on what properties and relations enter into the functional states that are identified with mental states. We have seen that hard-core functionalism requires the internal state (or states) that mediate between external inputs and external outputs to be physical states. And let us grant that the input and output states themselves, the states that serve to "anchor" functional roles, will be physical too (sensory inputs and behavioral outputs). What else enters into functional states? There are causal/nomological relations that connect anchoring inputs to internal mediators (mediators to other mediators) and mediators to anchoring outputs. In addition, there are logical functions (quantification, conjunction) and there is the quasi-logical relation of property instantiation. We can conclude, then,

[4] I intend this remark about multiple realizability as merely reporting the received view. I myself think it is a myth that hard-core functionalism allows for multiple realization; see section 5.

that the functional states in question are constituted of physical states together with so called "topic neutral" properties and relations. The term 'topic neutral' was used by Smart (1962, p. 60) to refer, roughly, to entities that are neither physical nor mental, at least not by definition. (Causal/nomological relations are included among the topic neutral relations so as not to beg any questions against dualism by definition.)

The upshot of all this is that hard-core functionalism is not hard-core physicalist reductionism. It reduces mental states to physical-cum-topic-neutral states. This kind of weak-minded physicalism will have to suffice for hard-core functionalists. They will point out that topic neutral properties, albeit not purely physical, are after all not mental either: there is no foothold for dualism here.

2. Kim Functionalism

Hard-core functionalism can be characterized by the following three points: (*i*) it identifies mental states with functional states; (*ii*) it holds that mental states are realized by physical states and that they may be (and probably are) multiply realized by physical states; and (*iii*) it is not, strictly speaking, a form of physicalist reductionism, at best, it is a form of extended physicalist reductionism. Kim opposes hard-core functionalism on all three points: (*i*) he does not acknowledge the existence of those mental states that hard-core functionalism wants to identify with functional states; (*ii*) he identifies the mental states he does acknowledge with first-order physical states, hence, his functionalism does not allow for multiple realization of mental states; (*iii*) his functionalism is a straightforward form of physicalist reductionism with respect to the mental states he acknowledges. This contrast to hard-core functionalism falls out of the following feature of Kim's view: he proposes a form of physicalist functionalism that denies the existence of functional states.

Consider, again, the simplified correlation biconditional:

(C) $x \text{ is in pain} \leftrightarrow (\exists\phi)(x \text{ has } \phi \;\&\; \mathrm{I}_{[\text{damage}]}\rightarrow\phi\rightarrow\mathrm{O}_{[\text{wincing}]})$.

Kim holds that there is no property, hence, no functional state expressed by the right hand side of this correlation. Since Kim also holds that the biconditional is true, he holds that there is no property expressed by the left-hand side of the correlation either: there is no such mental *state* as being in pain. Nevertheless, according to Kim, this does not mean that the correlation cannot be upgraded to something that is at least in the neighborhood of a type-identity thesis. But the upgraded thesis will concern *concepts* rather than states or properties; it will concern the concept of pain and the concept of having some (first-order) property that plays the functional role characteristic of pain. Whereas hard-core functionalism identifies a mental state (or property) with a functional state (a second-order property), Kim's functionalism equates the concept of a mental state M with a functional concept:

(KF) The concept of M ≈ the concept of having some (first-order) property ϕ, such that being in I causes one to have ϕ and having ϕ causes one to go into O.

I deliberately use the vaguer 'equate' ('≈') rather than 'identifies', because I am not sure whether Kim really wants to commit himself to a strict identity between mental-state concept and functional concept.[5]

Kim asks us to embrace a form of eliminativism. He proposes to eliminate the across-the-board applicable mental states that hard-core functionalism wants to identify with functional states. Our general concepts of pain, belief, desire, and so on are just that: they are just our concepts; they do not directly mirror any states or properties present in nature. But Kim's eliminativism does not reach any further than that. For he does hold that there are genuine mental states present in nature, namely mental states restricted to types of organisms—to species or, alternatively, to physical-

[5] For Kim's conceptual version of functionalism, see 1992, pp. 330-335; 1996, pp. 120-122; and this volume, pp. [25-30].

structure types. (The precise nature of the restriction is for science to find out. I shall assume that the restriction to species is good enough for present purposes.) Instead of pain, we have human pain, dog pain, martian pain, etc. Each such species-specific mental state is identical with some first-order physical state. Consequently, species-specific mental states are not multiply realizable. But each species-specific mental state can be said to realize a corresponding unrestricted mental-state *concept* in virtue of the fact that it is identical with some first-order physical state that realizes the unrestricted concept (e.g., human pain, which is identical with brain state C_{17}, realizes the concept of pain). It is the mental-state concept, rather than the mental state, that is multiply realized. Instead of one global reduction of pain, we get many *local reductions* of species-specific pain states. No global reduction is called for, because unrestricted pain has been eliminated.[6]

It may seem somewhat harsh to describe Kim as an eliminativist with regards to non-species-specific, unrestricted pain. After all, Kim does acknowledge that there is the unrestricted pain *concept*. Notice, however, for Kim this unrestricted concept does not enjoy the same status as the various species-specific pain states. The species-specific states, being identical with genuine physical states, are natural kinds; they are objectively "out there". The unrestricted pain concept, on the other hand, comprises our conception of what is objectively out there.[7]

The crucial ingredient in Kim's position is his distinction between properties, states, and natural kinds on the one hand, and concepts on the other hand. Kim (1992, p. 333) indicates that properties, states, and kinds are a matter of ontology, whereas concepts are a matter of epistemology and, one might add, practical interests. Concepts reflect our attempts to gain epistemic access to what is out there and our need to communicate about what is out there (cf. this volume, [p. 28f.]). If we adopt Kim's "epistemic interest" approach to concepts, we should not be surprised to find that our

[6] For Kim's local reductionism, see 1989, pp. 271-275; 1992, pp. 327-330; and 1996, pp. 233-236.

[7] See Kim 1992, p. 334, for his commitment to limited eliminativism.

concepts will often fail to “carve nature at its joints”, that is, they will often fail to mirror faithfully real-world states and properties. I assume that this should not be taken as the expression of some global skepticism to the effect that our concepts inevitably deceive us about the true structure of reality. Instead, it should be taken as the recognition that our concepts do reflect reality but often in a somewhat oblique manner. They constitute our perspective on nature’s joints, which means that it may sometimes take (scientific and/or philosophical) work to disentangle the contributions made by reality from the contributions made by us. On Kim’s view, our unrestricted mental-state concepts would be a case in point; they carve nature more along our needs than along its joints: species-specific (or physical-structure specific) mental concepts can do a better job; they stand for natural kinds.

How would a Kim functionalist go about defending his view vis-à-vis a hard core functionalist? Consider an example that Kim himself has used in a closely related context (cf. 1992, p. 319f.). Putnam has told us that the term ‘jade’ does not refer to a single mineral kind; instead, it subsumes two minerals with rather different physical makeup, jadeite and nephrite. Assuming this to be true, what are we to say about jade? The hard-core functionalist response would be that there are three kinds here, jadeite, nephrite, and jade; the latter is a functional kind that is realized by jadeite and nephrite. A Kim functionalist would disagree. His response would be that there are only two kinds here: jade is not a kind, it is merely a second-order functional concept. A disagreement of this nature is certainly not easy to resolve. But the “epistemic interest” approach to concepts points at least into the general direction in which a resolution should be sought. We should look at the functional role that enters into the definition of jade. Does it specify a feature common to minerals of a certain kind or is it more plausibly interpreted as specifying a feature common to *our* ways of interacting with certain minerals that are hard to distinguish on qualitative grounds? If one can make the case that plausible input/output conditions of the functional role characteristic of jade will specify epistemically tractable, publicly available conditions for the warranted applicability of the term ‘jade’, then one has at least the beginning of a case for regarding jade as a functional concept rather than a natural kind. However things actually stand with jade, the prospects for making an analogous case concerning mentality look fairly promising. After all, the functional roles that enter into functionalist definitions of the mental are anchored in external sensory inputs and external behavioral outputs. That is, they do indeed seem to trace rather closely the publicly and/or scientifically available evidence base for ascribing mentality to organisms. Although this line of reasoning needs to be worked out in more detail, it does seem to provide at least the beginning of a case for Kim’s view that functional definitions of pain, belief, desire, etc., define functional concepts rather than natural kinds.[8]

To summarize Kim’s position: All mental states are species specific, and each species-specific mental state is identical with a first-order physical state. There are no unrestricted, species-unspecific mental states, only unrestricted, species-unspecific mental-state concepts. And these concepts are (conceptually equivalent to) second-order functional concepts that are realized by species-specific mental states, that is, by first-order physical states. Kim’s functionalism is the “old” Feigl-Smart identity theory applied to all the mental states there are on Kim’s count and enriched with a twist of conceptual functionalism.

3. Three Research Projects for Kim Functionalism

I will now try to identify three general issues that arise specifically in connection with the conceptual aspect of Kim’s version of functionalism. I believe that upon closer examination at least two of them will quickly turn into problem areas. First, the position I referred to as hard-core functionalism is primarily motivated by multiple-realizability arguments. These arguments attempt to show that mental states can be

[8] These considerations are an attempt to develop some remarks made by Kim in 1992, p. 333; 1996, pp. 120ff.; and in this volume, [pp. 28f.].

realized by different first-order physical states. Since Kim is committed to the thesis that only mental-state concepts (but not mental states) can have multiple realizations, he is committed to the view that multiple-realizability arguments have been misapplied. That is, he is committed to the view that earlier functionalists have been confused (including Putnam who, given his other views, should have known better) and that their arguments can and must be reconstrued as pertaining to concepts rather than mental states if they are to be taken seriously. I have not checked into this, but my guess is that no serious obstacles stand in the way of such a conceptualizing reinterpretation of multiple-realizability arguments.

Second, as far as I can see, Kim has not stated explicitly what kind of equivalence relation he thinks obtains between an unrestricted mental-state concept (say, pain) and its "defining" functional concept (e.g., the concept of having some property that is caused by tissue damage and causes wincing and ...). In other words, Kim has not said what he takes '≈' in (KF) to mean: Is it identity or some weaker relation of conceptual equivalence? Kim has indicated that concepts are to be thought of along semantic lines. They are in the same ball park as predicates, meanings, ideas, Fregean senses, and synonymy classes of predicates.[9] If concepts are, roughly, meanings, does that mean that we should interpret the symbol '≈' in (KF), and the double arrow in (C) for that matter, as indicating that the flanking expressions are literally synonymous? This strikes me as a bold thesis. The expression on the right-hand side of (C) contains quantification over properties and it contains the idea of causal relations; its meaning is substantially richer (more sophisticated) than the

[9] See Kim 1992, p. 333; 1996, p. 120ff.; and this volume, [p. 28]. In one respect, Fregean senses (*Sinne*) seem especially apt. For we are used to thinking of them as "modes of presentation" of referents. And the notion of a mode of presentation, like Kim's notion of a concept, is largely driven by epistemic concerns. On the other hand, Frege's notion of sense is often associated with his thesis that sense determines reference, i.e., that expressions with the same sense must have the same referent. Kim, along with many other philosophers, will want to cancel this associated thesis. If I understand him correctly, Kim wants to hold that the unrestricted term 'pain' expresses the concept/sense *pain* and refers to any property that realizes that concept. Since different species-specific properties can realize the concept of pain, the term `pain' can have different referents without being ambiguous.

meaning of 'pain'. A relation of conceptual equivalence that commits one to regard the expressions corresponding to the concepts as synonymous is too demanding. But are there any relations that can do the job of conceptual equivalence without being committed to synonymy of corresponding expressions? Notice that philosophers who found themselves in similar quandaries have typically been able to fall back on the plausible view that the expressions they wanted to equate, though not synonymous, nevertheless refer to the same property or natural kind (viz., "Water = H_2O"). By the very nature of his thesis, Kim cannot take this rout.

One often distinguishes between "psychofunctionalism" and "common-sense functionalism". The distinction concerns the "source" for the specification of the functional roles that enter into functional definitions. According to psychofunctionalism, the functional role associated with pain must derive from the best (available) psychological theory of pain. According to common-sense functionalism, the functional role associated with pain derives from common-sense platitudes involving 'pain'. ("Pain causes one to wince", may be one such platitude.) Proponents of hard-core functionalism tend to prefer the best-psychological-theory approach. On the other hand, proponents of views similar to Kim's (Armstrong, Lewis) always prefer the common-sense approach. The reason for the latter preference is fairly obvious. The thesis that functional definitions define *our* mental concepts is utterly indefensible, unless these definitions proceed in terms of common-sense platitudes. Kim (1996, p. 110) suggests that psychofunctionalism and common-sense functionalism need not compete with each other: psychofunctionalism could be seen as defining a (the best available) scientific concept of unrestricted pain, whereas common-sense functionalism could be seen as defining the common-sense concept of unrestricted pain. The first part of this suggestion seems comparatively unproblematic. The strictly scientific concept of pain (if there is such a thing) could indeed be identified with whatever functional concept a (the best available) psychological theory has to offer. I am, however, troubled by the second part of Kim's suggestion because the point I made above remains essentially the same. While it is not downright absurd to identify the common-sense concept of pain with a common-sense functional concept (as it would be if it were identified with a psychofunctional concept), it is still very bold to maintain that the ordinary expression 'pain' is literally synonymous with the expression offered by common-sense functionalism. And if concept equivalence is not committed to synonymy of expressions, what relation is it?

The third issue I want to raise concerns possible dualistic leftovers. Whatever we take concepts to be, *having* or *grasping a concept* and *applying a concept* are surely mental phenomena. In particular, grasping an unrestricted mental-state concept like pain is itself a mental phenomenon. How will Kim's functionalism handle such mental phenomena? This question hints at a potentially awkward dilemma. On the one hand, Kim may want to treat grasping the concept of pain along the lines of his treatment of pain. If so, he will have to say that grasping the concept of pain is not a genuine mental state either; it is itself just a concept that is realized in various first-order physical states, i.e., in various species-specific mental states of grasping the concept of pain. This response will raise the bothersome question what kind of physical states these realizers could be. Moreover, it will set off a regress of ever more involved concepts that need to be accounted for, since Kim will now be faced with *the concept of grasping the concept of pain*, and so on. On the other hand, Kim may want to maintain that grasping the concept of pain is not a concept, that it is a genuine mental state that can be identified with a genuine functional state (and not just with a functional concept). If so, one should seriously wonder which argument X will allow him to take this option in this case. Whatever argument X turns out to be, why is X not equally applicable to pain itself? If X establishes that grasping the concept of pain is a genuine mental state, then X should equally establish that being in pain is a genuine mental state after all. But that would mean that Kim's functionalism would undermine itself. I think that, if Kim functionalism is to be a general view about the mental that avoids undermining itself, then Kim will have to take the first option. He will have to construe the grasping of a mental-state concept as being itself a functional concept rather than a functional state, and so on. And this would seem to indicate that there will inevitably be some concepts left that cannot be handled by the Kim functionalist version of physicalism: there will be dualistic leftovers.

4. Kim's Arguments for Kim Functionalism

So far I have not considered any of the arguments by which Kim arrives at his version of functionalism. His most recent paper, "The Mind-Body Problem: Taking Stock after Forty Years" (this volume), contains two such arguments. The first one appears at a point where Kim has just given a sketch of functionalism and begins to prepare the transition from the hard-core interpretation of functionalism to his conceptual interpretation. He remarks on how functionalism provides an explanation of the supervenience thesis:

> The mental supervenes on the physical because every mental property is a second-order functional property with physical realizers. And we have an explanation of mental-physical correlations: Why is it that whenever *P* is realized in a system, *s*, it also has mental property *M*? Because having *M* consists in having a property with causal specification *D*, and, in systems like *s*, *P* is the property (or one of the properties) meeting specification *D*. For systems like *s*, then, having *M consist in* having P. It isn't that when certain systems instantiate *P*, mental property *M* magically emerges ... It is rather that having *M*, for these systems, *is* just having *P*. (This volume, [p. 22])

This argument, if taken literally, verges on being incoherent. Kim's words seem to imply that M, the second-order functional property of having some property ϕ such that I→ϕ→O, *is identical with* the property P, if P realizes M in organism s_1, that is, if P is such that s_1 has P and I→P→O. But surely, this can't be right. The identity between the second-order property M and the first-order property P can't be contingent on P's realizing M—Can property identity be contingent on anything? The second-order functional property simply isn't identical with P, no matter what. Kim, of course, points this out himself a few pages later (this volume, [p. 27]). It turns out that the conclusion he really wants us to draw from his argument is that there is no such second-order functional property M. (For, otherwise, wouldn't it have to be identical with its realizer P?) That is, the conclusion he really wants us to draw is that functional expressions express concepts and not properties.

However, the argument does not establish this conclusion. If being A logically entails being B, then one can say, albeit somewhat misleadingly, that bringing about B "consists in" bringing about A. After all, all one has to do to bring about B is to bring about A, logic will do the rest. But this use of "consists in" should not be confused with identity. For if B does not entail A, then being B is not identical with being A. The hard-core functionalist will make just this type of response. According to hard core-functionalism, mind-body supervenience is explained by realization and *entailment* and not by realization and identity. The mental supervenes on the physical because s's having P_1, where P_1 occupies the functional role $I \rightarrow \phi \rightarrow O$, entails that s has M, the second-order functional property of having some property ϕ such that $I \rightarrow \phi \rightarrow O$. Since having M does not entail having P_1 (and does not entail having P_2, and so on), there will be no pressure to identify it (absurdly) with its realizer. Hence, there will be no pressure to discard functional properties in favor of functional concepts. Kim needs another argument.

Kim's second argument is somewhat harder to evaluate:

> For something to have second-order property *M* is for it to have some first-order property or other meeting a certain specification. Say, there are three such first-order properties, P_1, P_2, P_3. For something to have *M*, then, is for it to have P_1 or have P_2 or have P_3. Here there is a disjunctive proposition, or fact, that the object has one or another of the three first-order properties; that is exactly what the fact that it has *M* amounts to. There is no need here to think of *M* itself as a property in its own right—not even as a disjunctive property with the *P*s as disjuncts. By quantifying over properties we cannot create new properties any more than by quantifying over individuals we can create new individuals. (This volume, [p. 27f.])

The argument makes two distinct moves. Assume that M and the three first-order properties are as specified by Kim, while s is some organism. First move: the fact that s has the *alleged* property M = the fact that s has P_1 or s has P_2 or s has P_3. Second move: the latter fact is a disjunctive fact, in particular, it is not the fact that s has (P_1 or P_2 or P_3); there is no need to assume that there is the disjunctive property of having (P_1 or P_2 or P_3) at all. Therefore, having M is not (need not be acknowledged to be) a genuine property making up the fact that s has M.

Kim supports the first move of his argument, the identification of M with its expansion, by the claim that we cannot create new properties by quantifying over old ones: “That would be sheer magic”, he says (this volume, [p. 27]). However, this remark will not move the hard-core functionalist. After all, the hard-core functionalist maintains that he is describing properties already out there and not creating them. To further support his claim, Kim points out that, if someone murdered Jones, and the candidates are Smith, Gonzales, and Wang, then there is no fourth person (the murderer) over and above Smith, Gonzales, and Wang (this volume, [p. 28]). But here Kim doesn’t seem to get the analogy quite right. For the first move of his argument relies on the idea that the fact that s has M should be identified with the disjunctive fact that s has P_1 or s has P_2 or s has P_3. So what we should ask is whether the fact that someone murdered Jones should be identified with the disjunctive fact that Smith murdered Jones or Gonzales murdered Jones or Wang murdered Jones. And this is a murky question. Maybe the following consideration speaks against the proposed identification. Kim maintains that under the circumstances the fact that someone murdered Jones amounts to the fact that at least one out of Smith, Gonzales, and Wang murdered Jones. But the fact that someone murdered Jones does not seem to be that sensitive to the circumstances. Rather, it amounts to the fact that at least one *out of all the persons there are* murdered Jones. And that is a different fact than the one proposed by Kim. Kim’s fact entails but is not entailed by my fact. I have to admit, however, that I am a bit worried here. For if we have to be careful not to confuse properties with concepts, we should probably also try not to confuse facts with thoughts or sentential meanings. Unfortunately, I find it hard to tell whether my counter proposal is based on facts about facts or on facts about meanings.

Let us see what happens if we waive any doubts one might have about Kim's second argument. Let us assume its conclusion: M facts are disjunctive facts and do not involve M as a genuine property. Does this thesis seriously threaten hard-core functionalism? So far, it looks like s's having M is still a genuine (if disjunctive) fact—a slice of what is really "out there". This would seem quite sufficient for hard-core functionalism; the idea that M is in addition also a genuine property may be, as one says, "negotiable". But Kim makes a supplementary move in order to show that s's having M is not a genuine fact either:

> Suppose that in this particular case, *x* has *M* in virtue of having P_2, in which case the truth-maker of "*x* has *M*" is the fact that *x* has P_2. There is no further fact of the matter to the fact that *x* has *M* over and above the fact that *x* has P_2. (This volume, [p. 29])

But what is this relation of *truth-making*? The fact that s has P_2 does entail the fact that s has M. But can truth-making be entailment? I do not think so. Take the fact that s has F to be the conjunctive fact that s has F_1 and s has F_2 and s has F_3, and so on, for all properties of s. If truth-making were entailment, such a conjunctive fact would make true the fact that s has F_1. We could then use an argument analogous to Kim's argument to show (absurdly) that there is no such fact as s's having F_1. So truth-making cannot be entailment. What is it then?

In short, I find the arguments Kim offers in "The Mind-Body Problem" nconclusive. The first argument does not work. The second argument contains a problematic but hard to evaluate premise, and its completion makes use of the notion of truth-making which is not sufficiently spelled out. So far, the hard-core functionalist thesis that there are unrestricted mental states M that are identical with second-order functional states has not been refuted. But in some earlier work Kim has produced another argument, one that bypasses most of the issues raised above. This further argument relies on the "Causal Inheritance Principle":

> [Causal Inheritance Principle] If *M* is instantiated on a given occasion by being realized by *P*, then the causal powers of *this instance of M* are identical with (perhaps a subset of) the causal powers of *P*. (1993, p. 355)

Kim maintains that to deny this principle is to believe that a property could have excess causal powers over and above the causal powers of its physical realizers, it is to believe in magic: "Carburetors can have no causal powers beyond those of the physical structures that serve as one; and an individual carburetor's causal powers must be exactly those of the particular physical device in which it is realized (if for no other reason than the simple fact that this physical device *is* the carburetor" (1996, p. 118). The Causal Inheritance Principle alone, however, will not do all the work that Kim needs to have done. For the principle refers only to the causal powers of *instantiations* of the property M, whereas what is needed is a principle about the causal powers of being M *in general*. I think Kim should add a *principle of composition* to the effect that the causal powers of the property of being M are the causal powers of its (possible) instantiations taken together.[10]

The argument from causal inheritance is not yet complete. Kim wants to reach the conclusion that being M is not a genuine mental state or property. And to reach this conclusion, he invokes *Alexander's dictum*: *To be real is to have causal powers*.[11] This

[10] In 1992, p. 326, Kim points out that the Causal Inheritance Principle does not directly apply to the causal powers of being M *in general*. But he does not mention anything like a principle of composition in the argument he proceeds to give there. I am not quite sure how his argument can work without such a principle. (Maybe the somewhat different argument in 1996, pp. 118ff. is meant to repair this.) Moreover, it seems that composition is needed in any case. The Causal Inheritance Principle "condenses" the causal powers of instantiations of M to the causal powers of being P *in general* rather than to the causal powers of P's instantiations. If the causal powers of being P were not composed of the causal powers of the (possible) instantiations of P, a gap would be left from which causal powers could magically emerge.

[11] Kim 1992, p. 348; and 1993, p. 355f. The *dictum* is named after the emergentist Samuel Alexander.

is still not quite enough though. For M does have causal powers. According to composition and inheritance, its causal powers are just the causal powers of its (possible) instances, that is, of its (possible) realizers. To get the conclusion he wants, Kim needs to invoke a strengthened version of Alexander's dictum, which I will call *Alexander's kind-dictum*: To be real is to have causal powers; *and to be a real kind (state, property) is to have a nomologically unified set of causal powers*. This dictum says that genuine kinds or properties cannot be nomologically heterogeneous as causal powers; genuine kinds or properties collect things insofar as they have similar causal powers and play similar roles in causal laws.[12]

The beauty of Alexander's kind-dictum lies in the fact that it traps the hard-core functionalist in his own devices. Hard-core functionalism is motivated by the thesis that mental states can be realized in physical properties that are very different and it is committed to the thesis that these realizing properties are very different *because* they are radically heterogeneous as causal powers. Otherwise, there would be no obstacle to regarding the disjunction of the realizing properties as a natural kind in its own right. But this is precisely what functionalism does not want to do. If the disjunction of the physical properties that realize a mental state M were itself a genuine property, then M could simply be identified with this disjunction and functionalism would be superfluous. Therefore, hard-core functionalists themselves have to be committed to Alexander's kind-dictum. And this means that being in pain (i.e., having some property with the functional role characteristic of pain) cannot be a genuine mental state. For its causal powers are just the causal powers of its nomologically heterogeneous realizers. Being in pain can at best be a concept, namely the concept of having some property that occupies the relevant functional role. (A slightly modified version of this argument will show that s's being in pain cannot be a genuine fact.)

I find this argument extremely impressive. This is not to say that one can't

[12] I think Kim's arguments in 1992, p. 326f., and 1996, p. 118ff., make implicit use of Alexander's kind-dictum.

have serious qualms about Alexander's *dicta* or about the Causal Inheritance Principle. But one should remember that we have been engaged in an in-house debate among two forms of physicalist functionalism, a debate between Kim functionalism and hard-core functionalism. And it is difficult to see which of the principles a hard-core functionalist could reject consistent with his goals of being a physicalist, a functionalist, and being hard-core.

5. Contra Ramseified Functionalism

Can hard-core functionalism avoid Kim's argument from causal inheritance? Not if it is construed in the way presented above. However, this does not support Kim's conceptualized functionalism because the basic framework presupposed by both versions of functionalism—the hard-core version as well as Kim's version—does not work to begin with. Remember the point of functionalism is to provide an account of mentality that allows for multiple realizations. Consider again the simplified functionalist correlation biconditional concerning pain (which I have simplified even further):

(C) x is in pain $\leftrightarrow (\exists\phi)(\phi x \ \& \ (\mathrm{I} \rightarrow \phi \rightarrow \mathrm{O}))$.

Does this correlation allow for pain to be multiply realized? Upon inspection it turns out that the answer to this question must be negative.

The condition that specifies the functional role of ϕ, $\mathrm{I}\rightarrow\phi\rightarrow\mathrm{O}$, says that being in input state I causes one to go into ϕ, and being in ϕ causes one to go into output state O. This formulation is supposed to abbreviate a system of causal laws about organisms that exhibit the phenomenon of mentality; it is supposed to abbreviate a psychological theory (common-sense or scientific) that is true of such organisms. Notice that the functional role is specified in terms of property causation. And statements about property causation are equivalent to or entail (it doesn't matter for now) corresponding nomologically necessary universal generalizations. For instance, the statement that having F causes having G can be spelled out in terms of the universal statement that, for every *y*, if *y* has F then *y* has G, which has to be understood as a nomological necessity. This indicates that a slightly expanded version of (C) would have to take the following form:

(C_1) $\quad x$ is in pain $\leftrightarrow (\exists\phi)(\varphi x \;\&\; (\forall y)(\mathrm{I}y \rightarrow \phi y \rightarrow \mathrm{O}y))$.

It is easy to see that this correlation precludes the multiple realizability of pain. For it entails that an organism is in pain, only if there is one (at least one) physical property that occupies the functional role characteristic of pain *in all organisms y*. And this gets the quantifiers the wrong way round. It entails that the functional role cannot be occupied by different physical properties in different organisms. The problem is obviously independent from the question of whether one takes functionalism to be a view about functional states or about functional concepts. Hard-core functionalism as well as Kim functionalism, since they are both based on (C), are equally untenable because they do not allow for multiple realization.

What went wrong? The fault lies with the widely accepted Ramsey-Lewis method of developing functionalism via the "Ramseification" of a general theory.[13] To oversimplify drastically, assume the theory to be Ramseified has the general form of '$(\forall y)(\mathrm{F}y \rightarrow \mathrm{G}y)$' where 'F' stands for a theoretical term. To Ramseify, one replaces the predicate constant 'F' by a predicate variable, 'ϕ', and existentially quantifies the result. This yields the second-order existential generalization '$(\exists\phi)(\forall y)(\phi y \rightarrow \mathrm{G}y)$'. To arrive at functionalism, we have to proceed analogously, or so we are told. First we

[13] See Lewis 1983. Kim explicitly develops functionalism along the lines of the Ramsey-Lewis method in 1996, pp. 105-107; and this volume, [pp. 16ff]. His other writings seem to presuppose the same understanding of functionalism.

have to take a psychological (common-sense or scientific) theory about pain, say, "Tissue damage causes pain which causes wincing", and then we have to Ramseify over its theoretical predicate, i.e., 'pain'. But putting it this way hides the fact that the theory is really a universal generalization with respect to organisms y. It has the form '$(\forall y)(\mathrm{I}y \rightarrow \mathrm{PAIN}y \rightarrow \mathrm{O}y)$'. If we Ramseify this theory, we will arrive (after a trivial step) at (C_1). And this creates the problem because it gets the quantifiers the wrong way round. Moreover, it seems clear that there is no simple way of repairing this problem through alterations of the embedded theory (the functional role): no amount of finagling 'R' will make a claim of the form '$(\exists x)(\forall y)(x\mathrm{R}y)$' come out as a claim of the form '$(\forall x)(\exists y)(x\mathrm{R}y)$'.

6. Hard-Core Structural Functionalism?

Can this problem be repaired without destroying functionalism? I think so, but the Ramsey-Lewis method has to be discarded. What is needed is a formulation in which the physical-property quantifier '$(\exists\phi)$' and the organism-quantifier '$(\forall y)$' have traded places. Instead of saying that to be in pain is to have some (first-order) physical property ϕ that occupies the role characteristic of pain in every organism, we need to say that to be in pain is to be one of those organisms in which some physical property ϕ occupies the characteristic role and to have ϕ. But what is meant by "one of *those* organisms in which some ϕ occupies the characteristic role"? To cash this out, we need to say that organisms fall under physico-biological structure types, *ST*, that determine whether a certain functional role can be realized in an organism by some first-order physical property ϕ.[14] Putting this together results in the thesis that to be in pain is to belong to some structure type *ST*, such that, in every organism belonging to *ST* the functional role characteristic of pain is occupied by some physical property ϕ, and to

[14] The use I make here of structure types is influenced by Kim's formulation of a structure-restricted realization thesis in 1992, p. 313.

have φ. In general:

$$(C_3) \quad x \text{ has } M \leftrightarrow \{\exists ST\}\{STx \,\&\, (\forall y)(STy \rightarrow (\exists \phi)(\phi x \,\&\, (Iy \rightarrow \phi y \rightarrow Oy)))\}.$$

The first-order physical property P_1 realizes M in organism *x* just in case (i) *x* belongs to one of the structure types *ST*, such that for every organism *y* belonging to that structure type, $(\exists\phi)(Iy \rightarrow \phi y \rightarrow Oy)$; (ii) *x* has P_1; and (iii) P_1 is one of the physical states ϕ that can occupy the functional role $(Iy \rightarrow \phi y \rightarrow Oy)$.

Unfortunately, this proposal is rather cumbersome. I have not yet found a way of spelling it out in more illuminating terms. Nevertheless, I believe that—after some scrutiny—one can discern that the new formulation allows for a mental state M to be realized by different physical properties in different organisms, as long as the organisms belong to one of the organism types in which some physical property occupies the functional role that is characteristic of M. If this proposal works out, then the above correlation can be "upgraded" to a theoretical identification giving rise to an alternative (and even more abstract) form of hard-core functionalism: The property of being in mental state M *is* the property of belonging to some physical structure type that determines the functional role characteristic of M. (Of course, the 'M' in the definiens has to be "discharged" by actually spelling out the characteristic functional role.)

I suggest that this alternative form of hard-core functionalism might be able to withstand Kim's argument from causal inheritance. If so, the main reason for switching over to Kim's conceptualized version of functionalism would be removed.[15]

[15] Shoemaker (1981, pp. 264-267) suggests a functionalist correlation significantly simpler than (C_3) which would also seem to avoid the multiple-realizability problem described in the previous section. According to his account, a functionalist definition of a mental state M would be based on a correlation of the form: "x has M $\leftrightarrow (\exists\phi)(\phi x \,\&\, (Ix \rightarrow \phi x \rightarrow Ox))$". My preference for (C_3) is based on two considerations. First, Shoemaker's proposal is just as vulnerable to the causal-inheritance argument as the original version of functionalism, hence, it does not remove the pressure for switching over to Kim's

Here is a very rough sketch of how (C_3) might avoid the causal-inheritance argument. Assume that the first-order physical property P_1 realizes mental state M in organism s of type ST_1. In virtue of what does this instance of M have the causal powers it has? Say s's having M causes some physical event e to occur. In this instance, then, it is s's having P_1 that causes e. However, and this is the crucial part, *P_1's causing e to occur* is, in this instance, itself caused by P_1's occurring within a structure of type ST_1. (If P_1 had not occurred within structure ST_1, then it would not have caused e at this occasion, unless of course P_1 had occurred within one of the other structures that determine the functional role characteristic of M.) The idea would be that M's causal powers derive from the physical structures embodied in organisms having M. These physical structures function as enabling causes allowing the physical realizers of M to cause the physical events they cause. Moreover, the structures are not heterogeneous; they are unified by the following feature: they determine every organism that embodies one of them to have one and the same characteristic functional role realized by some physical state. The structures are unified by the single functional role they determine. And this is what unifies the causal powers of mental states.[16]

conceptualized version of functionalism. Second, it is hard to see how Shoemaker's proposal can be made to fit into the standard nomological account of causality. The second conjunct in the right-hand side of Shoemaker's correlation is not a universal generalization, nor can it be regarded as an instance of a universal generalization for otherwise the proposal would again succumb to the multiple-realizability problem. This means that we cannot read it as a causal claim, at least not on the standard nomological account of causality. Notice that the second conjunct in the right-hand side of (C_3) is a universal generalization, hence, it can be read as a causal claim.

[16] An earlier version of this paper was delivered as a response to Kim's Perspective Lecture at the University of Notre Dame in November, 1995. Thanks to Paddy Blanchette, Terry Horgan, Jaegwon Kim, Michael Kremer, and Leopold Stubenberg for helpful comments.

References

Kim, J. (1989): “The Myth of Nonreductive Materialism”, reprinted in Kim (1993a), pp. 265-284.

—— (1992): “Multiple Realization and the Metaphysics of Reduction”, reprinted in Kim (1993a), pp. 309-336.

—— (1993): “The Nonreductivist’s Trouble With Mental Causation”, reprinted in Kim (1993a), pp. 336-357.

—— (1993a): *Supervenience and Mind. Selected Philosophical Essays*, Cambridge, Cambridge University Press.

—— (1996): *Philosophy of Mind*, Boulder, Westview Press.

—— (This volume): “The Mind-Body Problem: Taking Stock after Forty Years”.

Lewis, D. (1970): “How to Define Theoretical Terms”, in *The Journal of Philosophy*, 67, pp. 427-446.

Shoemaker, S. (1981): “Some Varieties of Functionalism”, reprinted in *Identity, Cause, and Mind. Philosophical Essay*, Cambridge, Cambridge University Press 1984, pp. 261-286.

Smart, J. J. C. (1962): “Sensations and Brain Processes”, reprinted in C. V. Borst, ed., *The Mind/Brain Identity Theory*, St. Martins Press, New York 1970, pp. 52-66.