
Abstract: Jaegwon Kim has argued that a popular line of reasoning from multiple realizability to the falsity of psychophysical reductionism has the unwanted effect of ruling out psychological laws altogether. Setting aside the issue of reductionism, I consider three ways one might try to show that multiple realizability poses a problem for psychological laws. I conclude that none succeed outright, and the last enjoys only a conditional success: if certain sorts of explanations are to be had, then there is an epistemic peril in positing psychological laws, but it is entirely unclear whether those sorts of explanations are to be had.

Multiple Realizability and Psychological Laws:
Evaluating Kim's Challenge

D. Gene Witmer
University of Florida

1. The significance of multiple realizability

Suppose, as many believe, that (at least some) mental properties are multiply realizable. What is the significance of such multiple realizability? A standard answer is that psychophysical reductionism is rendered untenable. The argument for this may be briefly characterized as follows. Suppose R, R, etc. are all the physical realizers of pain. Then, if psychology is to reduce to physics, a completion of the following must be a legitimate bridge principle:

\[(x) \text{ (x has pain if and only if x has R or x has R or ...)}\]

The antireductionist argument turns on the claim that there is something wrong with supposing this to be an apt bridge principle. Intuitively, the open sentence "x has R or x has R or ..." seems to collect various instances which fail to be sufficiently similar to each other to be grouped under a common natural category. As a result, it cannot pick out a property suitable for citation in a bridge principle, and reductionism fails. The property fails to be the sort of thing that can figure in a law of nature and (hence) is not a genuine property or "kind" in the first place.

In his "Multiple Realizability and the Metaphysics of Reduction" (1992; "MRMR" hereafter), Jaegwon Kim claims that this familiar argument proves too much: if we accept its premises, we will block psychophysical reduction only at the cost of being forced to accept the further, disturbing conclusion that mental properties themselves cannot appear in laws and are not genuine "kinds" in the first place. Whatever is wrong with the property expressed by the disjunctive predicate infects, so to speak, the mental property itself. This is Kim's challenge: if there is something about that "disjunctive" property that prevents it from appearing in laws of nature, how could mental properties themselves appear in laws of nature?

While Kim's challenge is directed at those philosophers who use multiple realizability as an argument against reductionism, those unsympathetic to that argument should still be concerned with Kim's challenge. For that challenge does not depend on that context. It arises for anyone who is inclined to accept both that there are psychological laws and that mental properties are multiply realized. If (like myself) you think the original antireductionist argument is fundamentally misguided, you may still need to find an answer to Kim's challenge.

In this paper I develop and evaluate Kim's challenge independently of issues about intertheoretic reduction. For the sake of continuity with Kim's paper, I will focus on pain, supposing that it is multiply realizable by virtue of being a functional property—that is, the property of having some property that plays such and such a causal role. I will also continue to suppose that R, R, ... are all the possible physical realizers of pain. The alleged property picked out by the open sentence "x has R or x has R or ..." will be named the "physical-realizer sum (the PR-sum) of pain." My focus will be on developing Kim's remarks into arguments for the thesis that something about the PR-sum of pain forces us to the conclusion that laws involving pain are impossible, unlikely, or in some way epistemically problematic.

Multiple Realizability and Psychological Laws
I develop three lines of argument. The first aims to establish that pain is not projectible—that is, capable of figuring in generalizations which are confirmed by their positive instances. If this is right, then even if there are laws involving pain, we cannot confirm their existence in the usual fashion. While this argument—which I call simply the Unprojectibility Argument—is the most natural one to extract from MRMR, it is, I think, quite unpersuasive.

The second line of argument turns on Kim's "causal inheritance principle," according to which any causal powers possessed by a functional property must be inherited by it from those possessed by its realizing physical properties. This principle implies that any laws involving a functional property must be explicable by virtue of being derivable from laws governing its physical realizers. If, then, there is some reason to think that such derivability is ruled out, there is reason to think there are no laws governing pain. What I call the Inexplicability Argument maintains that the multiple realizability of pain stands in the way of such derivability. I suggest on Kim's behalf a number of ways one might argue that such derivability is blocked, but none of them, I argue, succeeds.

The third argument, finally, combines considerations of projectibility with the causal inheritance principle in a way which helps explain why Kim's worries about multiple realizability have the intuitive force that they do. This last argument attempts to show that if we project a generalization to cases realized differently from those already observed, we will be committing ourselves to an inexcusable proliferation of independent explanations. While the previous two arguments are, I think, unsuccessful, this one—the Explanatory Multiplication Argument—enjoys a limited success. It does not (I will argue) establish that such projections must be unwarranted; it does, however, establish a conditional thesis, namely, that if explanations of a certain sort are forthcoming, then we ought not to project to cases realized differently from those already observed. It is entirely unclear whether those explanations are forthcoming, however. As a result, the ultimate verdict of this paper must be that Kim's challenge gives us no reason to despair of psychological laws involving multiply realizable properties.

2. The Unprojectibility Argument

2.1 The argument introduced

In rehearsing the standard antireductionist argument, Kim focuses on the claim that the PR-sum of pain is unfit for laws and complains that typical attempts to defend this claim are unilluminating:

What we now need is an argument for this claim; to dismiss such disjunctions as "wildly disjunctive" or "heterogeneous and unsystematic" is to label a problem, not to offer a diagnosis of it. (MRMR, 318)

He then turns to offer his own diagnosis of it, suggesting that the key notion is that of projectibility. His challenge to defenders of the antireductionist argument is then to explain why pain itself should be thought projectible. So here is one simple argument to be found in MRMR:

The Unprojectibility Argument

1. The PR-sum of pain is unprojectible.
2. If the PR-sum of pain is unprojectible, then pain itself is unprojectible. Hence, pain is unprojectible.

If we conclude that pain is unprojectible, we may go on to conclude that there can be no laws involving pain; whether we do so depends on whether we think projectibility is actually a prerequisite for appearing in a law of nature. Setting that question aside, however, it would be in itself disheartening for those optimistic about psychology to learn that we cannot confirm generalizations about pain by observing positive instances thereof.

Those suspicious of Kim's argument but enthusiastic about the original antireductionist argument are likely to focus on the second premise, which we may call the "linking thesis," as it links the unprojectibility of the PR-sum of pain with the unprojectibility of pain itself. I argue
below that the linking thesis is actually rather plausible. But the first premise, as I shall argue in §2.3, is utterly unsupported.

2.2 The linking thesis

The simplest way to defend the linking thesis would be by identifying pain with its PR-sum. So long as projectibility is a genuine property subject to Leibniz’s Law, such identity would guarantee the linking thesis. Kim seems inclined to pursue this route at times:

> We cannot hide the disjunctive character of pain behind the second-order expression, "the property of having a property with specification X". Thus, on the construal of mental properties as second-order properties, mental properties will in general turn out to be disjunctions of their physical realization bases. (MRMR, 323-4)

The argument here seems to rely on a more general principle to the effect that, for any second-order property, that property is identical with the disjunction of all its possible realizers. If this principle is to have any plausibility, the disjunction of realizers must include every possible realizer, no matter how distant from the actual world. This point raises an immediate potential problem for the identification Kim wants to draw, for there may exist possible nonphysical realizers of pain. In that case, the PR-sum of pain cannot be identified with pain, for the simple reason that it fails to be necessarily coextensive with pain. Say that Q₁, Q₂, etc., are all those possible realizers of pain that are not themselves physical, and let the “comprehensive-realizer sum (CR-sum) of pain” be the alleged property picked out by the open sentence “x has R₁ or x has R₂...or x has Q₁ or x has Q₂ or...” Kim’s claim in the above passage, then, is plausible only if the CR-sum of pain is identical with the PR-sum of pain.

Whether that identity is sufficient to warrant identifying pain with its PR-sum is another, difficult question I will not tackle here. Let us just say that there is at least room for resisting the identification of pain with its PR-sum. Even so, there is reason to buy the linking thesis. That premise may be supported by breaking it into two subsidiary links thus:

(2a) If the PR-sum of pain is unprojectible, then so is the CR-sum of pain.

(2b) If the CR-sum of pain is unprojectible, then so is pain itself.

While there may be good reason to resist identifying the CR-sum of pain with pain itself, their coextension is metaphysically necessary. As a result, any correlation between pain and some feature F entails a correlation between the CR-sum of pain and F as well. Plausibly, then, if a generalization about pain is confirmed, the corresponding generalization about its CR-sum is likewise confirmed. The projectibility of pain thereby implies the projectibility of its CR-sum; so 2b seems very plausible.

One might object here to my presumption that two necessarily equivalent hypotheses must also be alike with regards to whether they are confirmed by a particular case. If this presumption is false, then the Unprojectibility Argument can be blocked in this fashion. Indeed, if confirmation is understood in an appropriately intensional fashion, we could probably allow that pain is identical with its PR-sum and still maintain that the PR-sum is unprojectible while pain is projectible. Since I do not endorse the Unprojectibility Argument, this is fine with me. Nonetheless, I do not want to rest my rejection of the argument on that consideration, for two reasons. First, I want to be able to engage those who are not inclined to take confirmation to be understood in such a fashion. Second, it seems to me that talk of projectibility is in the first place motivated by considerations hostile to such an understanding. Projectibility is presumably a very worldly matter, not dependent on the way in which a property is expressed. (Grueness is presumably unprojectible regardless of what predicate is used to express it, for instance.) If that is right, then there is good reason to think that my presumption is correct.

So let us suppose 2b is correct and turn to 2a. Why think that if the PR-sum is unprojectible, then so is the CR-sum? Whatever makes a property unprojectible is presumably a matter of having some instances that differ from each other in some special way; to take the classic example, certain instances of grue differ from each other in ways that make the projection...
of grueness problematic. If that is what suffices for unprojectibility, however, 2a is guaranteed. If the PR-sum of pain is unprojectible, then there are instances of the PR-sum of pain which differ in that significant way. Those instances will also be instances of the CR-sum of pain; hence, the CR-sum of pain will also have instances that differ in the way that makes for unprojectibility. In light of these considerations, it is not, I think, very promising to reject Kim’s challenge by denying the linking thesis.

2.3 Disjunctiveness, “nothing in common,” and unprojectibility

A better tack is to deny the initial claim that the PR-sum of pain is unprojectible. What accounts for the widespread intuition to the contrary? The first thing we need to bear in mind is that our grasp of the PR-sum of pain is exceedingly limited; it is not as if we have already produced an uncontroversial functional analysis of pain and identified the disjunction of physical realizers. No casual examination of the PR-sum can do any work here.

Nor is there any shortcut to establishing the unprojectibility of the PR-sum of pain by way of focusing on the notion of disjunctiveness. If it seems obvious to you that a generalization involving the PR-sum of pain is unconfirmable because it involves a “disjunctive” property, you are likely relying on the following two claims: (i) properties can be classified as “disjunctive” or not independently of intuitions regarding projectibility; and (ii) those which are disjunctive are not projectible. But there is no good reason to believe both of these together. If “disjunctive” just means “can be expressed using a disjunction of predicates,” then (ii) is obviously false. If (ii) is taken for granted, on the other hand, we have no clue as to how to decide whether a property is “disjunctive” unless we already have come to some conclusion regarding its projectibility. We might put the point thus: While Kim worries that the predicate “pain” hides the disjunctive character of the property, one might just as well worry that the disjunctive predicate used to express the same property imposes a false impression of unprojectibility.

A more promising approach to our question is to focus on the idea that there is “nothing in common” between the various physical realizers of pain. If we can establish that claim, then (perhaps) we can go on to conclude that the PR-sum of pain is unprojectible. But establishing that initial claim is the hard part. Indeed, it seems out of the question, for there is obviously something in common to all the realizers of pain—for they all manage, after all, to realize pain.

The present point can be made precise by invoking a distinction introduced by Sydney Shoemaker (1981) that seems to me insufficiently appreciated in the literature on functionalism. He distinguishes between “core” and ”total” realizers in the following way. Let us suppose the functional definition of ”pain” is this:

\[ x \text{ is in pain} = \text{there is some property } P \text{ such that } x \text{ has } P \text{ and } x \text{'s having } P \text{ meets condition } \psi. \]

Suppose an individual has neurological property \( N_i \), and that individual’s having of \( N_i \) meets condition \( \psi \). In what is perhaps the most common usage of “realize,” \( N_i \) realizes pain on that occasion. This is what Shoemaker means by core realization. For every core realizer, however, there is a total realizer—intuitively, the core realizer conjoined with the condition of actually meeting \( \psi \). In this case, then, the total realizer may be expressed thus:

Having \( N_i \) and being such that one’s having \( N_i \) meets \( \psi \).

When I introduced talk of the PR-sum of pain, I did not specify (in the main text) what sort of realizer I meant to collect in the disjunction. Kim’s discussion in MRRM does not make any explicit choice between them. I take it, however, that it is the disjunction of physical total realizers that is relevant, since Kim presumes that a realizer of pain is sufficient for pain, and only total realizers are guaranteed to be sufficient. For the remainder of this paper, then, by ”realizer” I shall mean ”total realizer,” unless otherwise specified.

In light of this clarification, we can make more precise what it is all the disjunctions in the PR-sum of pain must have in common. If \( N_1, N_2, \) and so on are all the physical properties capable of core realizing pain, then the list of total physical realizers (\( R, R, \ldots \)) can be expressed thus:
Having N₁ and being such that one's having N₁ meets condition \( \psi \).

Having N₂ and being such that one's having N₂ meets condition \( \psi \).

... 

This point has been previously noted by Louise Antony and Joseph Levine (1997: 90). Each total realizer is bound to have something in common with every other total realizer: each is definable in a way that includes the requirement that some instance of a property meets the condition \( \psi \). So it is just false to say that there is nothing in common between the various disjuncts of the PR-sum.

Still, this consideration is hardly conclusive. Depending on what condition \( \psi \) looks like, the commonality above may seem too trifling to banish the threat of unprojectibility. One might, in fact, argue as follows. Functionalists typically envisage a condition that is relational: the core realizer must be related to other properties in such-and-such ways. If condition \( \psi \) is relational, one may be tempted to take that as conclusive evidence of the unprojectibility of the PR-sum.

I suspect this line of thought has much to do with Kim's conviction that functional properties are not projectible. In MRMR, he acknowledges that instances of pain will have this much in common: each core realizing property will be such as to satisfy the definitive condition \( \psi \). But he emphasizes that this commonality does not imply that different instances of pain have anything intrinsic in common:

That, one might say, is the whole point of functionalism: the functionalist, especially one who believes in [multiple realizability], would not, and should not, look for something common to all pains over and above [their satisfying \( \psi \)] (the heart of functionalism, one might say, is the belief that mental states have no "intrinsic essence"). (MRMR, 332)

The argument here seems to be this. First, if a category C is such that its instances have nothing in common with each other except relational features, C is not projectible. Second, the instances of the PR-sum of pain have nothing in common with each other except relational features. Hence, the PR-sum of pain is not projectible.

One might protest that the second premise is unsupported. The fact that the definitive condition \( \psi \) is relational does not guarantee that the various physical total realizers have nothing intrinsic in common. But suppose we accept the second premise; the argument still fails, as the first premise is simply too restrictive. Consider these generalizations:

Anything written after brainstorming is better than something written without any brainstorming.

All products of such-and-such company sell well.

If you get a good night's sleep, you will be much more productive the next day.

These may or may not qualify as potential laws, but they are certainly confirmable by their positive instances. Yet the category of (for instance) being a product of such-and-such company is an excellent candidate for one whose instances have nothing intrinsic in common.

One might object that I am begging the question here, since such generalizations subsume cases that exhibit the same kind of variety as we find with multiply realizable properties. The various papers written after brainstorming could, for example, be extremely different from each other. But I take it that, so far as intuition goes, these are obviously confirmable. (At the very least they do not display the kind of startling inability to be confirmed famously exhibited by "grue.") Argument might overturn this prima facie verdict, but the burden is on he who denies their confirmability. If the advocate of the Unprojectibility Argument complains about my appeal to intuition, I can only ask whether his own judgement that the PR-sum of pain is unprojectible is based on anything better.

So far, it seems we cannot find any support for the claim that the PR-sum of pain is unprojectible. Indeed, it is tempting to respond to these attempts with a Moore-style argument. It is a Moorean fact that we have good reason to believe, on the basis of a number of positive
instances, generalizations about pain. Given the linking thesis, then, we can infer by modus tollens that the PR-sum of pain is projectible.

I take this argument to have considerable weight. We at least have some epistemic intuitions regarding pain, which can be counted as evidence. But we have, really, no intuitions about the PR-sum of pain, for the simple reason that we have not actually specified any such property. If we actually had some properties in hand, some with which we were familiar, our projectibility intuitions about realizer-sums might be worth counting. As it is, it seems the best argument to be had for taking any stand on the projectibility of the PR-sum of pain is the Moorean argument for it.

3. The Inexplicability Argument

3.1 The "metaphysical mystery" and the causal inheritance principle

In his critical commentary on Kim's challenge, Fodor (1997) suggests that "what's really bugging Kim" is "a metaphysical mystery about functionalism" (159):

Damn near everything we know about the world suggests that unimaginably complicated to-ings and fro-ings of bits and pieces at the extreme microlevel manage somehow to converge on stable macrolevel properties. . . . [T]he 'somehow' really is entirely mysterious, and my guess is that that is what is bugging Kim. He just doesn't see why there should be (how there could be) macrolevel regularities at all in a world where, by common consent, macrolevel stabilities have to supervene on a buzzing, blooming confusion of microlevel interactions.... (160-161; emphasis original)

Fodor's diagnosis here is, I think, quite on target. Although it is easy to get bogged down in the matter of projectibility, the appeal of Kim's challenge is best traced instead to the thought that, first, any law about pain must be due to laws about its physical realizers, but second, that the variety of physical realizers somehow stands in the way of such explanation.

The "metaphysical mystery" can be brought into focus by looking at Kim's "causal inheritance principle." In MRMR, Kim formulates the principle as follows:

If mental property M is realized in a system at t in virtue of physical realization base P, the causal powers of this instance of M are identical with the causal powers of P. (MRMR, 326)

How exactly is this principle relevant to the appearance of multiply realizable properties in laws? I take it the link is roughly this: laws must reflect causal powers, so the laws involving a multiply realizable property F must, given the causal inheritance principle, be determined by the causal powers (hence the laws) governing the realizers of F. In other words, if a multiply realizable property F is governed by a law L, its being so governed must be entirely due to the laws governing the various realizers of F.

This interpretation is confirmed by a more recent discussion in which Kim introduces the causal inheritance principle by first discussing Block's claim (1990) that dormitivity can be a cause of cancer. Dormitivity is here defined as the second-order property of having some property that induces sleep. Kim writes:

Block in fact claims that the dormitivity of Seconal, in spite of its admitted impotence to cause sleep, can have effects of other kinds, for it is possible, he says that there is a law to the effect that dormitivity of a pill is causally sufficient for the ingester's getting cancer. However, as is evident, this could be the case only if each specific chemical realizer of dormitivity caused cancer; it is difficult to see how dormitivity as such, independently of its realizers, can cause cancer or anything else. (Kim 1998, 54)

The causal inheritance principle is then explicitly introduced as a generalization of this claim. If there is a law about a second-order property M, that law must be determined by the
laws governing the physical realizers of M. More precisely, if L is a law governing M, then the fact that L is a law must be explicable as a result of those realizer laws; it must be capable of being explained by being derived from those laws.\textsuperscript{v} I take it, then, that the causal inheritance principle is relevant to our question by virtue of motivating what I will call the Explicability Requirement (ER):

\begin{quote}
(ER) For any law L which involves pain, L can (in principle) be explained by being derived from the laws governing the physical realizers of pain.
\end{quote}

We can generalize the motivation for ER as follows. Functional properties are themselves not basic but depend for their instantiation on other properties; as a result, any laws governing such properties must themselves depend on other laws—the laws governing the more basic properties. Since we are presuming pain is a functional property, any law governing pain must itself be explicable as a result of more basic laws—in particular, the laws governing the physical realizers of pain. The kind of explanation wanted is a derivation: we want to explain how the law governing pain is entailed by the laws governing its physical realizers. Whether we can ever produce such an explanation is beside the point; our confidence that there are such nonbasic laws depends on our confidence that such an explanatory derivation could, in principle, be given.

But should we be so confident? The Inexplicability Argument that I propose on Kim’s behalf is inspired by the thought that it is extremely unlikely that such an explanatory derivation could be given. How could a single pain law be derived from such a wild variety of laws governing different physical realizers? If such derivation seems hopelessly implausible, one may well take this to be a good reason to suppose that there are no laws governing pain in the first place. Consider, then, this argument:

**The Inexplicability Argument**

1. For any law L which involves pain, L can (in principle) be explained by being derived from the laws governing the physical realizers of pain.  
   (ER)
2. It is very unlikely that any law involving pain could be so derived.

Hence, we should conclude that there are no laws involving pain.

It seems to me that ER is secure. To tackle the second premise, however, we need to do some more digging. What is supposed to make it unlikely that any law involving pain could be derived from the laws governing its physical realizers?

Given our assumption that pain is a functional property, there seems to be no principled bar to such derivation. The functional definition of pain will underwrite metaphysically necessary bridge principles of the form “if x has R, then x is in pain.” Such principles make it clear how the laws governing the various physical realizers of pain could entail nomologically necessary generalizations about pain itself. The mere possibility of deriving a pain law from the laws governing the physical realizers of pain is, then, unproblematic.\textsuperscript{vi} There might, however, be reason to think such derivability is exceedingly unlikely; in particular, the fact of multiple realizability might give one reason to doubt that pain laws are so derivable. It is this worry that constitutes the Inexplicability Argument.

I explore two ways one might develop this worry. The first turns on the claim that the physical realizers differ in that they are not governed by any common physical laws; this lack of commonality may seem to render psychological laws inexplicable. The second suggestion focuses not so much on the differences between the various realizers but on the sheer number of them. The lack of constraints on what could realize pain may be thought to result in far too many distinct realizers for an explanatory derivation to be available. Neither suggestion, I argue, establishes that the Explicability Requirement is unlikely to be met; so neither gives us good reason to think that there are no laws governing pain.

3.2 The lack of common governing laws
Here is one way one might get worried about the derivability of pain laws from the laws governing the physical realizers of pain. Recall the Unprojectibility Argument; the primary intuition there was that the PR-sum of pain is unprojectible because the various instances of that property are in some sense importantly different from each other. What is striking about their variety is that they seem likely not to be the sorts of properties one would group together if one were hoping to develop a physical theory. As Antony and Levine put it, "from the first-level point of view, the collection of realizers is arbitrary" (1997, 90). In other words, there seems no motivation, within physical science, to group these together as a kind liable to be subject to laws. If the various physical realizers of pain were governed by common physical laws, then they would be apt to be grouped together as a single category in the physical sciences; the PR-sum would thereby be salient from a physical point of view.

This thought suggests a route to supporting the second premise of the Inexplicability Argument. Let the following be an alleged psychological law, where Q is some physical property:

(A) Pain events cause Q-events.

Recall Kim’s remarks on the possibility of a law dictating that dormitivity cause cancer; there, he suggested that such a law would only be possible if, for each physical realizer of dormitivity, there was a law dictating that it cause cancer. Similarly (A) implies the existence of several physical laws.

R-events cause Q-events.
R-events cause Q-events.
R-events cause Q-events.
....

I take it that this convergence of different physical properties on a common physical effect would be a striking fact from the perspective of physical science. It is exactly this sort of commonality that seems missing from the collection of various physical realizers of pain, and it is the lack of such commonality that seems responsible for the persistent intuition the PR-sum of pain must be unprojectible.

Let us say that there is a common physical law governing properties R and R just in case there exist physical laws L and L such that L mentions R, L mentions R, and L and L are otherwise equivalent. More precisely, where L and L are both physical laws, if a statement of L can be transformed into a statement of L simply by replacing each mention of R with a mention of R, and vice versa, then R and R are governed by a common physical law. With this terminology in hand, we can say what allegedly makes it unlikely that (A) could be derived from the laws governing the physical realizers of pain. Consider the following "No Common Physical Laws" (NCPL) thesis:

(NCPL) There is no common physical law connecting all of the physical total realizers of pain with the effects typically attributed to pain.

Given NCPL, we can produce a simple argument against the existence of laws involving pain. Suppose that there is a law involving pain. Given the explicability requirement ER, there is some common physical law governing all the physical total realizers of pain. But there is no such common physical law, at least according to NCPL. So the initial supposition that there is a law involving pain must be false.

The argument is invalid. It does not in fact follow from ER and the assumption that there exists a law involving pain that there is a common physical law governing all the physical total realizers of pain. While ER may imply that there must be a common law governing the physical total realizers of pain, it does not imply that there must be a common physical law governing those realizers.

To see this, bear in mind that likely psychological laws will not relate a single mental property to a single physical property. More likely is a law in which other multiply realizable properties are involved. So, for instance, (B) is not unlikely:
(B) Pain-events cause increases of vigilance.

I here presume that vigilance is multiply realized as well as pain. Let us suppose the physical total realizers of vigilance are \( V_1, V_2, \ldots \) and so on. What physical laws might suffice for the derivability of (B) from laws about the physical realizers of pain and vigilance-increasing? Here is one possible scenario:

\[
\begin{align*}
R_1 &\text{-} \text{events cause } V_{35} \text{-} \text{events.} \\
R_2 &\text{-} \text{events cause } V_{35} \text{-} \text{events.} \\
R_3 &\text{-} \text{events cause } V_{28} \text{-} \text{events.} \\
&\text{\ldots}
\end{align*}
\]

Whereas both \( R_1 \)-events and \( R_2 \)-events cause \( V_{35} \)-events, the other physical realizers of pain cause distinct physical events. So this set of laws is hardly guaranteed to be salient for physical science. This set of laws could, nonetheless, provide an explanatory derivation of the psychological law (B).

The intuition that I invoked earlier held that the different realizers did not seem salient as a class from the perspective of physics. Someone interested only in studying physical phenomena would not find any interesting laws governing all of those realizers. When we consider the physical variations it seems to us unlikely that any one working on physical theory would be interested in grouping together the physical realizers in a single class.

But we can now see that this intuition, and hence NCPL, may be respected while making room for the derivability of pain laws by the laws governing the physical realizers of pain. It may, indeed, be necessary that there be some common law governing those realizers for there to be such an entailment. For instance, (B) might require that \( R_1, R_2, \text{and so on} \) are all governed by a law dictating that they cause vigilance-increase events. But that law is not likely to be of interest for physical theory, as the type "vigilance-increase" is not salient for the development of physical theory. The arbitrariness of the PR-\( \text{sum of pain from the point of view of physical theory is no threat to the explicability of pain laws as resulting from laws governing those physical realizers.}

3.3 The lack of constraints and the sheer number of realizers

My concern in criticizing the argument from NCPL was to show that the truth of NCPL is compatible with the derivability of pain laws from laws governing the physical realizers of pain. But one might also have responded by wondering why we should believe NCPL in the first place: what knowledge do we have, anyway, of the character of the physical realizers of pain? After all, as I stressed earlier, it’s not as if we have any actual specification of these realizers to work with. Whatever is driving the intuitions is not the specific character of properties named as likely realizers. What makes NCPL strike philosophers as plausible is rather, I suggest, a more general intuition concerning the ease with which a property can realize a functional property. The easier it is to play a given functional role, the more physical realizers there are; the more physical realizers there are, the less likely they will have anything in common of interest to physical science.

Once this thought is out on the table, however, a distinct approach to substantiating the Inexplicability Argument suggests itself. Instead of relying on some claim about the different physical character of the various disjuncts of pain’s PR-\( \text{sum}, \) we might rely on a claim about the sheer number of such realizers. If in fact there is a distinctive lack of constraints on what could realize pain, then an enormous number of distinct realizers is possible; it then becomes unreasonable to suppose that some derivation capturing all of them is forthcoming.

Think again about the dormitivity example. If it is a law that dormitivity causes cancer, then for every physical total realizer \( R \) of dormitivity, it is a law that \( R \) causes cancer. Given how dormitivity is defined, however, it seems there will be an enormous number of distinct physical realizers of dormitivity; it then seems absurd to suppose that there exists, for each one of these, a physical law dictating that it cause cancer. In brief, the worry is that a law involving a multiply realizable property would need to be sustained by the existence of a huge number of independent physical laws, where the latter eventuality strikes us as unlikely.

The worry here is best appreciated by thinking first in terms of \( \text{core} \) realizers. Assuming, again, that pain is the property of having some property \( P \) such that one’s having \( P \) meets
condition $\psi$, if $\psi$ concerns only the role played by the core realizing property, it seems that any number of different physical properties could core realize pain. For every distinct core realizer of pain there is a distinct total realizer of pain. Too many of these, and it seems derivability is unlikely.$^{i i i}$

There are two thoughts here which we should distinguish. The first is the (admittedly vague) thesis that the number of distinct physical realizers of pain is enormous. The second is the claim that the greater the number of distinct physical realizers of pain, the less likely it is that the laws involving pain are derivable in the way demanded by ER. If the "enormous" number invoked in the first thesis is sufficiently high, we can conclude that it is extremely unlikely that any law involving pain is derivable from the laws governing its physical realizers. While this argument may seem hard to evaluate, given its reliance on some very fuzzy gestures towards what is an "enormous number" and the like, we can still defuse the argument quite a bit by taking a closer look at both theses.

In his recent evaluation of Kim's challenge, Block has in effect attacked the first thesis. He writes:

In Walt Disney movies, teacups think and talk, but in the real world, anything that can do those things needs more structure than a teacup. We might call this the Disney Principle: that laws of nature impose constraints on ways of making something that satisfies a certain description. There may be many ways of making such a thing, but not just any old structure will do. It is easy to be mesmerized by the vast variety of different possible realizations of a simple computational structure, say that of an and gate, which can be made of cats, mice and cheese as well as mechanical or electronic components. But the vast variety might be cut down to very few when the function involved is mental, like thinking, for example, and even when there are many realizations, laws of nature may impose impressive constraints. (Block 1997, 120)

Block's point is exactly right. It is easy to overlook the "Disney Principle" so long as one forgets how complex the functional roles definitive of mental states are likely to be on any remotely believable version of functionalism. If we evaluate the mental case by thinking about, say, dormitivity—a very simple example of a second-order property—we may be overwhelmed by the lack of constraints imposed by such second-order specifications. With dormitivity, the core realizer only needs to do one thing: induce sleep. That does not seem like the sort of thing that could engage the Disney Principle to much effect. The functional role of any given mental state is bound to be much more complex: a core realizer of pain will presumably need to be relied on to cause a wide variety of different sorts of effects in a wide variety of situations as well as be itself the reliable effect of a wide variety of different causes. Solving such an engineering problem is no easy task; the more complex the definitive role, the more constraints that are imposed by the Disney Principle. So it is hardly obvious that the number of distinct physical realizers of pain is bound to be enormous.

The second thesis—that the greater the number of realizers the less likely that pain laws can be derived—is also questionable. It relies on an oversimplified picture of psychological laws. Suppose we had an alleged pain law of the following form:

(C) Pains cause W-events.

To derive this law, it seems that every physical total realizer of pain must be such that it causes W-events. But no realistic law will be so simple. What makes (C) unrealistic is its failure to be sensitive to the importance of background factors. Compare it to (D):

(D) Pains in circumstances of type C cause W-events.

It is obvious that any plausible pain law will need to be qualified in this way. Pains suffered when unexpected have different effects than those suffered when anticipated; pains endured when one knows the cause have different effects than pains endured when one does not know the cause. Such qualifications are inevitable.
Such qualifications also make it easier to derive the pain laws from the laws governing the physical total realizers of pain. To derive a law concerning pains occurring in circumstances of type C we need only consider those physical realizers that can be instantiated in circumstances of type C. So it is inaccurate to say that every increase in the number of possible physical realizers makes it harder to derive the laws involving pain. Those laws will be limited in several ways in the first place.

The present point is related to the suggestion made by both Block (1997) and David Papineau (1993) that evolutionary explanations relieve the worry that multiple realizability renders problematic the explicable psychological laws. In brief, the idea is that the different realizers of pain typically end up conforming to similar patterns as a result of selective pressures; a deeper unity is to be discerned in the common explanation of why each of the different realizers obey that pattern. There are two distinct ways in which this suggestion may be employed. One way is to consider it relevant to the entailment of psychological laws by offering a plausible qualification of any proposed psychological laws. Those laws which we have reason to posit may be restricted in scope to cognitive systems which have been produced after a sufficiently lengthy and rich period of evolution. The other way to employ the suggestion here is more radical but is in keeping with certain philosophical approaches to mentality. Instead of qualifying the laws at issue, we might insist that mentality is essentially linked to natural selection, so that by definition any instance of pain must have had such a history. If we take this tack, we are, in effect, enriching the definitive condition \( \psi \) so that it contains the resources for explaining why typical instances of pain conform to the law in question. This way of employing the suggestion can be seen as an extension of the Disney Principle: if it is part of the definitional constraints on the realization of pain that it have such-and-such a history, then the Disney Principle kicks in, since there will be only so many ways in which something could have such a history.

4. The Explanatory Multiplication Argument

4.1 Projecting to differently explained cases

While the arguments examined so far are unsuccessful, I do not think Kim’s worries are entirely groundless. He is right to point to a certain epistemic peril when it comes to laws involving multiply realizable properties. The peril is made manifest in the following passage:

Consider a possible law: “Sharp pains administered at random intervals cause anxiety reactions”. Suppose this generalization has been well confirmed for humans. Should we expect on that basis that it will hold also for Martians whose psychology is implemented (we assume) by a vastly different physical mechanism? Not if we accept the Physical Realization thesis, fundamental to functionalism, that psychological regularities hold, to the extent that they do, in virtue of the causal-nomological regularities at the physical implementation level. The reason the law is true for humans is due to the way the human brain is “wired”; the Martians have a brain with a different wiring plan, and we certainly should not expect the regularity to hold for them just because it does for humans. (MRMR, 324)

The force of Kim’s intuition here is striking. How exactly should it be understood? The first point to be stressed is that the conclusion aimed at here need not be understood as the claim that pain is an unprojectible property, period. We need to distinguish the very strong claim that a property is incapable of appearing in a generalization that is confirmed by its positive instances from the claim that, in a certain kind of situation, a certain kind of generalization is not confirmed by its positive instances.18 The more modest claim is what is at issue here. Let us make that claim precise. We can call it the Limited Ban on Projection (LBP) thesis:

\[ \text{(LBP)} \quad \text{If (i) every pain so far observed has had feature } F, \text{ and (ii) no observed pain has been realized by } R, \text{ then those observations do not warrant us in generalizing to the claim that it is a law that all pains, including those instantiated by virtue of being realized by } R, \text{ have } F. \]
While LBP is not as radical as the claim that pain is simply unprojectible, it is hardly insignificant. The worry is not only about exotic others such as Martians, of course. If multiple realizability is as widespread as some functionalists think, then nearly any mind outside of the cases taken as the inductive base will be one that is analogous to a Martian mind. If the conclusion LBP is correct, and multiple realizability really is that widespread, then the prospects for psychology are indeed dim.

It is easy to share Kim’s intuition about the human/Martian case. But what underlies the intuition? I suggest that the rationale is best understood in terms of an Occam-style principle of not multiplying explanations beyond necessity. Given the explicable requirement ER, to posit the law extending to Martians is to commit ourselves to an explanation beyond that to which we are already committed, one for which we have no independent warrant. By accepting the law as applied only to humans, we are already committed to there being a derivation of that law by the laws governing the human realizers of pain; an extension of the law to Martians would require an additional derivation of the law as applied to Martians by the laws governing the Martian realizers of pain. It is, I suggest, our sense that extending the law to Martians would presumably commit us to such an additional explanation that prompts our intuition that we would have no warrant for so extending it.

The epistemic rule here at work might be slightly generalized as the following "Explanatory Conservation" (EC) rule:

(EC) Where M. and M. are two kinds of Ms, all observed Ms have been M.s and none M.s, and every observed M has had feature F, then, if, in order for "all Ms have F" to be a law there must be an explanation of the fact that all Ms have F which is independent of any explanation of the fact that all Ms have F, then we should not, on the basis of observing that every M. has F, posit the law that all Ms, including both M.s and M.s, have F.

EC is suggested by our intuitions here, but it would be good if we could say more in its defense. How exactly is it motivated by Occam’s Razor? I take it that Occam’s Razor is best understood as a "tie-breaker" principle. Given two theories T. and T. if T. incurs fewer commitments than T. then, if T. and T. are in all other epistemic respects equal, then we should adopt T. and drop T.. If we posit the law that all Ms, including both M.s and M.s, have F, then we thereby incur more commitments than by positing the more restricted law that all M.s have F.

Are the two posits otherwise equal in epistemic respects? To evaluate this question with respect to any actually posited law would be a complex task, of course, but EC abstracts away from other epistemic issues by limiting its conclusion to the claim that you should not posit the broader law on the basis of those observed instances. There may be something else about the broader law which would warrant our positing it, but insofar as EC goes—and, for that matter, our ban LBP—the claim is only about the significance of those observations for our epistemic position, not our epistemic position all things considered.

Given EC, we can argue for LBP in the following fashion. First, suppose that the antecedent of LBP is satisfied. That is, suppose that (i) every pain so far observed has had feature F; and (ii) none of the observed pains has been realized by R. Let "pain." be the kind of pain that is realized by R. Now the argument may be set out thus:

The Explanatory Multiplication Argument

1. For any law L which involves pain, L can (in principle) be explained by being derived from the laws governing the physical realizers of pain. (ER)
2. Since pain, and pain, are realized by distinct physical properties, the explanations alluded to in the first premise are independent of each other.
3. Where M. and M. are two kinds of Ms, all observed Ms have been M.s and none M.s, and every observed M has had feature F, then, if, in order for "all Ms have F" to be a law there must be an explanation of the fact
that all Ms have F which is independent of any explanation of the fact that all Ms have F, then we should not, on the basis of observing that every M has F, posit the law that all Ms, including both Ms and Ms, have F. (EC)

Hence, these observations do not warrant us in generalizing to the claim that it is a law that all pains, including those instantiated by virtue of being realized by R, have F.

Discharging the conditional supposition, we have our limited ban on projection LBP. Kim’s intuition about this sort of case seems to be vindicated.

4.2 Must we refrain from extending the law to Martians?

The Explanatory Multiplication Argument is, I think, the most important one to emerge from our discussion. But it does not succeed in establishing LBP outright. I want to agree that it may well turn out that we lack the warrant for extending psychological generalizations beyond cases realized in the same way in which the base cases are. But I want to insist, further, that whether this is the case turns on the actual character of the explanation of the conformity of observed cases to the generalization at issue, and it is not inevitable that this explanation will not extend to differently realized cases.

Let us approach this point by thinking again about the Martian example. For simplicity, suppose that there is a unique realizer R of pain in humans and a unique realizer R of pain in Martians. The basic idea of the argument on the table is that if we suppose that the R, cases of pain conform to the law as well, then we are supposing that, in addition to the one explanation of the conformity of the R cases to the alleged psychological law, there is another, independent explanation of the conformity of R cases to that law. But this supposition is not necessary. The dubious move in Kim’s reasoning occurs when he says:

"The reason the law is true for humans is due to the way the human brain is "wired"; the Martians have a brain with a different wiring plan, and we certainly should not expect the regularity to hold for them just because it does for humans.

(MRMR, 324)

"The reason the law is true for humans" refers to the explanation of the conformity of R to the law by deriving it from physical laws governing R. Kim’s claim is that this explanation cannot also explain why the law would hold for the Martians; he takes this, apparently, to be an implication of the fact that Martian pain is realized differently. In short, he assumes that the difference between R and R, rules out the possibility that whatever explains why R conforms to the law also explains why R, conforms to it. This assumption is just the second premise of the argument as formulated above.

But the assumption is hardly mandatory. There may be an explanation of the conformity of R, to the law which is rich enough to imply, further, that instances of R, will conform to it as well. How is this possible? Here is a simple model. Let the law under consideration be that pain-events cause W-events. Now, since we have observed that human cases conform to this law, the following must be a law:

(E) R-events cause W-events.

So far, so good. But now here is the mistaken line of thought: since (E) is the explanation of why human pains conform to the law, and (E) will not entail that Martian pains conform to it as well, to posit an explanation of the unrestricted law is to posit an independent explanation, to multiply explanations beyond necessity.

The unnecessary supposition here is that (E) is all the explanation to be had of why human pain conforms to the law. After all, (E) itself need not be the bottom line; it is quite possible—indeed, likely, given the complexity of the physical property R—that (E) is to be derived from other, more fundamental laws. Since R is a total realizer, it is complex; there is some physical property N such that R can be represented as the property of having N and being such that one’s having N meets the condition definitive of pain—condition ψ. Suppose N is a
property of sort S, and that it is a more fundamental law that any property of sort S, when that property also satisfies condition \( \psi \), causes a W-event. In that case, the core realizer embedded in \( R \), might also be of sort S, and the more fundamental explanation will extend to Martian pains as well.

The key point is that, whereas it seems \( (E) \) must be involved in the explanation offered, there is no reason to think that the explanation is limited to \( (E) \): a deeper explanation might be available which is more general in such a way as to explain why R-events conform to the law as well. There is of course no guarantee that this be the case, but my point is only that multiple realizability does not rule it out.

The upshot, then, is that, although LBP has not been shown to be true, it remains a live possibility that certain projections of psychological laws which seem warranted may turn out not to be. If a particular projection is one in which the antecedent of LBP is satisfied and any explanation of the conformity of the unobserved sorts of cases would be independent of the explanation of the conformity of the observed kinds of cases, then that projection is unwarranted; in that case, we have no reason to extend the generalization to the unobserved sorts of cases. Whether LBP legitimately applies to a particular projection, then, depends on the actual character of the explanations which, given ER, must exist.

Generalizations involving mental properties which may seem well-confirmed can be undermined by certain explanations that may be forthcoming if mental properties are indeed multiply realized. We are far from being in a position at the moment, however, to make even an educated guess as to whether such explanations are to be had. To have even the roughest of guesses we would need to have some idea, not only of the appropriate functional analyses of various psychological states, but also of how the conformity of observed instances to generalizations we want to make is in fact to be explained. And we are nowhere near knowing either of these.  

5. Conclusion

Kim’s challenge in his “Multiple Realizability and the Metaphysics of Reduction” is aimed at those who think multiple realizability serves to render psychophysical reductionism untenable. My goal in this paper has not been to defend that claim about multiple realizability, for I think that the antireductionist argument is flawed for reasons untouched by Kim’s discussion. My goal, rather, has been to consider whether multiple realizability somehow stands in the way of supposing that there are laws governing functional properties. Does the multiple realizability of, say, pain imply that generalizations about pain cannot be confirmed, or that it is exceedingly unlikely that there are laws governing pain, or that projections to newly realized cases are never warranted? I have argued that the first two of these alleged consequences can be firmly set aside; the third depends on exactly how the observed sorts of cases are in fact to be explained. I take it that Kim’s challenge has some intuitive force, that when we consider, for instance, the possibility of projecting to the Martian cases, we are struck with the possibility that such projection may be entirely unwarranted. My formulation of the Explanatory Multiplication Argument is meant to diagnose this intuitive force while rendering it possible for us to evaluate it critically. The result is that the intuition reflects the fact that it may, in fact, turn out that such projection would commit us to entirely novel, independent explanations in a way that is plainly unacceptable. But it also might not: it depends on the actual character of the explanation of conformity of observed sorts of cases to the law in question. Ultimately, then, we can explain why Kim’s challenge can strike us as having so much force without agreeing that multiple realizability must have such dire consequences.

REFERENCES

When a reprint is cited, all page references are to the reprint.


Multiple Realizability and Psychological Laws


For useful comments and discussion I wish to thank William Lycan, Sven Walter, and the members of my spring 2000 graduate seminar on physicalism.

1 How is it misguided? Here’s the ultra-brief story. The original antireductionist argument presumes that psychophysical reductionism requires bridge principles of the sort described in the main text, ones linking mental properties to a disjunction of their physical realizers. But this is a mistake. While psychophysical reductionism requires bridge principles, and while those plausibly need to be property identities, they need not link mental properties to those disjunctions. Multiple realizability is usually taken for granted only because of the prior presumption that mental properties can be defined as functional properties; as a result, those functional definitions can themselves serve as appropriate bridge principles.

11 For those sensitive to Shoemaker’s distinction between core and total realizers (1981), I stipulate that I mean R₁, R₂, ... to be total realizers of pain. Kim does not declare whether he means total realizers or core realizers by “realizers” in MRMR, but his discussion makes most sense on the supposition that he means total realizers. The core/total distinction is brought into the main text in §2.3.

111 While Kim worries both about whether multiply realizable properties are “genuine kinds” and about whether they can appear in laws of nature, I take it that the former question depends on the latter.

iv It is worth stressing that this complex property is itself defined in physical terms, since the condition ψ will itself be physical; it will be, anyway, if the functional definition given above is supposed to provide a sufficient condition for the existence of pain in a physicalist world.

v Not all explanations consist in derivations, of course; nor do all derivations amount to explanations. But in this sort of case, where what we want to explain by reference to another general law is itself a general law, derivation seems to me to be the heart of the desired explanation.

vi I want to note two complications here that are incidental to the main argument of this paper. (i) One might object here that establishing that a generalization about pain is nomologically necessary is not enough to establish that it is a law. If this is right, then even with those bridge principles it may be hard to see how the realizer laws could entail that the generalization about pain is a law. While I am doubtful that there is a significant distinction to be made here between the nomologically necessary and the lawful, if there is such a distinction then it is plausible that the Explicability Requirement may be weakened so as to require only that the realizer laws entail the nomological necessity of the generalization citing pain itself. (ii) A rather more interesting complication concerns the possible nonphysical realizers of pain. Suppose the law that is allegedly entailed by the laws governing the physical realizers of pain is the law that
pain events cause increases of vigilance. If this is the actual law, it apparently governs those instances of pain that are realized by the nonphysical realizers as well as physically realized cases. But there is no way that the laws governing the physical realizers will entail that the nonphysically realized instances will conform to the law. In response to this I think the right thing to say is that we have no reason to believe in any laws governing both physical and nonphysical realizers of pain. See footnote 10 for more on this.

vii This thesis could be weakened to imply only that there is no law subsuming even most or a majority of the physical total realizers of pain. Such a weakened version may be relevant if the law we want to derive is understood as having exceptions. The extension of my comments to such a weakened premise should be obvious.

viii Of course, only those physical realizers which are capable of being instantiated without violating the actual laws of physics are relevant here. Any physical total realizer of pain that cannot be thus instantiated cannot threaten the derivability of pain laws from the laws of physics. Still, this restriction may not seem to come to much.

ix Indeed, I doubt that any contingently instantiated property is unprojectible, period. Even the infamous case of grue fails to be unprojectible in this strong sense. Suppose that grue is defined thus: x is grue iff either x is examined before 2100 and green or not examined before 2100 and blue. Now suppose, further, that in the year 2200 a large quantity of emeralds not examined before 2100 have been unearthed, and every one of them has been (to our immense surprise) blue. In that case, the generalization "all emeralds are grue" has, I daresay, some inductive support. (In light of this point, the reader should note that it is really no surprise that I could not find any merit in the simple Unprojectibility Argument first discussed.)

x While we may be able to extend the law to Martians, we cannot justifiably extend it to all possible types of pain, including those realized by nonphysical properties. If pain is defined as a second-order property, then nonphysical properties could play the requisite role and thereby core realize pain. The laws governing the physical realizers of pain are certainly not going to be capable of explaining why nonphysical realizers of pain conform to the law in question. This limit on our ability to project psychological generalizations is, however, quite benign. Psychology is hardly deprived of its pretensions once it is forbidden from trying to draw conclusions about the psychological laws governing ghosts, angels, and spirits from observations of embodied minds.

xi See note 1.