

## International Phenomenological Society

---

Multiple Realization and the Metaphysics of Reduction

Author(s): Jaegwon Kim

Source: *Philosophy and Phenomenological Research*, Vol. 52, No. 1 (Mar., 1992), pp. 1-26

Published by: International Phenomenological Society

Stable URL: <http://www.jstor.org/stable/2107741>

Accessed: 10/11/2008 14:12

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=ips>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



*International Phenomenological Society* is collaborating with JSTOR to digitize, preserve and extend access to *Philosophy and Phenomenological Research*.

<http://www.jstor.org>

# Multiple Realization and the Metaphysics of Reduction

JAEGWON KIM  
*Brown University*

## I. Introduction

It is part of today's conventional wisdom in philosophy of mind that psychological states are "multiply realizable", and are in fact so realized, in a variety of structures and organisms. We are constantly reminded that any mental state, say pain, is capable of "realization", "instantiation", or "implementation" in widely diverse neural-biological structures in humans, felines, reptiles, mollusks, and perhaps other organisms further removed from us. Sometimes we are asked to contemplate the possibility that extraterrestrial creatures with a biochemistry radically different from the earthlings', or even electro-mechanical devices, can "realize the same psychology" that characterizes humans. This claim, to be called hereafter "the Multiple Realization Thesis" ("MR",<sup>1</sup> for short), is widely accepted by philosophers, especially those who are inclined to favor the functionalist line on mentality. I will not here dispute the truth of MR, although what I will say may prompt a reassessment of the considerations that have led to its nearly universal acceptance.

And there is an influential and virtually uncontested view about the philosophical significance of MR. This is the belief that MR refutes psychophysical reductionism once and for all. In particular, the classic psychoneural identity theory of Feigl and Smart, the so-called "type physicalism", is standardly thought to have been definitively dispatched by MR to the heap of obsolete philosophical theories of mind. At any rate, it is this claim, that MR proves the physical irreducibility of the mental, that will be the starting point of my discussion.

Evidently, the current popularity of antireductionist physicalism is owed, for the most part, to the influence of the MR-based antireductionist argument originally developed by Hilary Putnam and elaborated further by

---

<sup>1</sup> On occasion, "MR" will refer to the *phenomenon* of multiple realization rather than the *claim* that such a phenomenon exists; there should be no danger of confusion.

Jerry Fodor<sup>2</sup>—rather more so than to the “anomalist” argument associated with Donald Davidson.<sup>3</sup> For example, in their elegant paper on nonreductive physicalism,<sup>4</sup> Geoffrey Hellman and Frank Thompson motivate their project in the following way:

“Traditionally, physicalism has taken the form of reductionism—roughly, that all scientific terms can be given explicit definitions in physical terms. Of late there has been growing awareness, however, that reductionism is an unreasonably strong claim.”

But why is reductionism “unreasonably strong”? In a footnote Hellman and Thompson explain, citing Fodor’s “Special Sciences”:

“Doubts have arisen especially in connection with functional explanation in the higher-level sciences (psychology, linguistics, social theory, etc.). Functional predicates may be physically realizable in heterogeneous ways, so as to elude physical definition.”

And Ernest LePore and Barry Loewer tell us this:<sup>5</sup>

“It is practically received wisdom among philosophers of mind that psychological properties (including content properties) are not identical to neurophysiological or other physical properties. The relationship between psychological and neurophysiological properties is that the latter *realize* the former. Furthermore, a single psychological property might (in the sense of conceptual possibility) be realized by a large number, perhaps an infinitely many, of different physical properties and even by non-physical properties.”

They then go on to sketch the reason why MR, on their view, leads to the rejection of mind-body reduction:<sup>6</sup>

---

<sup>2</sup> Jerry Fodor, “Special Sciences, or the Disunity of Science as a Working Hypothesis” (hereafter, “Special Sciences”), *Synthese* 28 (1974): 97–115; reprinted in *Representations* (MIT Press: Cambridge, 1981), and as the introductory chapter in Fodor, *The Language of Thought* (New York: Crowell, 1975).

<sup>3</sup> Donald Davidson, “Mental Events” reprinted in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980).

<sup>4</sup> “Physicalism: Ontology, Determination, and Reduction”, *Journal of Philosophy* 72 (1975): 551–64. The two quotations below are from p. 551.

<sup>5</sup> “More on Making Mind Matter”, *Philosophical Topics* 17 (1989): 175–92. The quotation is from p. 179.

<sup>6</sup> “More on Making Mind Matter”, p. 180.

"If there are infinitely many physical (and perhaps nonphysical) properties which can realize  $F$  then  $F$  will not be reducible to a basic physical property. Even if  $F$  can only be realized by finitely many basic physical properties it might not be reducible to a basic physical property since the disjunction of these properties might not itself be a basic physical property (i.e., occur in a fundamental physical law). We will understand 'multiple realizability' as involving such irreducibility."

This antireductionist reading of MR continues to this day; in a recent paper, Ned Block writes:<sup>7</sup>

"Whatever the merits of physiological reductionism, it is not available to the cognitive science point of view assumed here. According to cognitive science, the essence of the mental is computational, and any computational state is 'multiply realizable' by physiological or electronic states that are not identical with one another, and so content cannot be identified with any one of them."

Considerations of these sorts have succeeded in persuading a large majority of philosophers of mind<sup>8</sup> to reject reductionism and type physicalism. The upshot of all this has been impressive: MR has not only ushered in "non-reductive physicalism" as the new orthodoxy on the mind-body problem, but in the process has put the very word "reductionism" in disrepute, making reductionisms of all stripes an easy target of disdain and curt dismissals.

I believe a reappraisal of MR is overdue. There is something right and instructive in the antireductionist claim based on MR and the basic argument in its support, but I believe that we have failed to follow out the implications of MR far enough, and have as a result failed to appreciate its full significance. One specific point that I will argue is this: the popular view that psychology constitutes an *autonomous special science*, a doctrine

---

<sup>7</sup> In "Can the Mind Change the World?", *Meaning and Method: Essays in Honor of Hilary Putnam*, ed. George Boolos (Cambridge University Press: Cambridge, 1990), p. 146.

<sup>8</sup> They include Richard Boyd, "Materialism Without Reductionism: What Physicalism Does Not Entail", in Block, *Readings in Philosophy of Psychology*, vol. 1; Block, in "Introduction: What is Functionalism?" in his anthology just cited, pp. 178–79; John Post, *The Faces of Existence* (Ithaca: Cornell University Press, 1987); Derk Pereboom and Hilary Kornblith, "The Metaphysics of Irreducibility" (forthcoming in *Philosophical Studies*). One philosopher who is not impressed by the received view of MR is David Lewis; see his "Review of Putnam" in Block, *Readings in Philosophy of Psychology*, vol. 1.

heavily promoted in the wake of the MR-inspired antireductionist dialectic, may in fact be inconsistent with the real implications of MR. Our discussion will show that MR, when combined with certain plausible metaphysical and methodological assumptions, leads to some surprising conclusions about the status of the mental and the nature of psychology as a science. I hope it will become clear that the fate of type physicalism is not among the more interesting consequences of MR.

## II. Multiple Realization

It was Putnam, in a paper published in 1967,<sup>9</sup> who first injected MR into debates on the mind-body problem. According to him, the classic reductive theories of mind presupposed the following naive picture of how psychological kinds (properties, event and state types, etc.) are correlated with physical kinds:

For each psychological kind *M* there is a unique physical (presumably, neurobiological) kind *P* that is *nomologically coextensive* with it (i.e., as a matter of law, any system instantiates *M* at *t* iff that system instantiates *P* at *t*).

(We may call this “the Correlation Thesis”.) So take pain: the Correlation Thesis has it that pain as an event kind has a neural substrate, perhaps as yet not fully and precisely identified, that, as a matter of law, always co-occur with it in all pain-capable organisms and structures. Here there is no mention of species or types of organisms or structures: the neural correlate of pain is invariant across biological species and structure types. In his 1967 paper, Putnam pointed out something that, in retrospect, seems all too obvious:<sup>10</sup>

“Consider what the brain-state theorist has to do to make good his claims. He has to specify a physical-chemical state such that any organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusc’s brain (octopuses are mollusca, and certainly feel pain), etc. At the same time, it

---

<sup>9</sup> Hilary Putnam, “Psychological Predicates”, in W. H. Capitan and D. D. Merrill, eds., *Art, Mind, and Religion* (Pittsburgh: University of Pittsburgh, 1967); reprinted with a new title, “The Nature of Mental States”, in Ned Block, ed., *Readings in Philosophy of Psychology*, vol. 1 (Cambridge: Harvard University Press, 1980).

<sup>10</sup> “The Nature of Mental States”, p. 228 (in the Block volume).

must not be a possible brain of any physically possible creature that cannot feel pain.”

Putnam went on to argue that the Correlation Thesis was *empirically false*. Later writers, however, have stressed the multiple realizability of the mental as a *conceptual* point: it is an a priori, conceptual fact about psychological properties that they are “second-order” physical properties, and that their specification does not include constraints on the manner of their physical implementation.<sup>11</sup> Many proponents of the functionalist account of psychological terms and properties hold such a view.

Thus, on the new, improved picture, the relationship between psychological and physical kinds is something like this: there is no single neural kind  $N$  that “realizes” pain, across all types of organisms or physical systems; rather, there is a multiplicity of neural-physical kinds,  $N_h, N_r, N_m, \dots$  such that  $N_h$  realizes pain in humans,  $N_r$  realizes pain in reptiles,  $N_m$  realizes pain in Martians, etc. Perhaps, biological species as standardly understood are too broad to yield unique physical-biological realization bases; the neural basis of pain could perhaps change even in a single organism over time. But the main point is clear: any system capable of psychological states (that is, any system that “has a psychology”) falls under some structure type  $T$  such that systems with structure  $T$  share the same physical base for each mental state-kind that they are capable of instantiating (we should regard this as relativized with respect to time to allow for the possibility that an individual may fall under different structure types at different times). Thus physical realization bases for mental states must be relativized to species or, better, physical structure-types. We thus have the following thesis:

If anything has mental property  $M$  at time  $t$ , there is some physical structure type  $T$  and physical property  $P$  such that it is a system of type  $T$  at  $t$  and has  $P$  at  $t$ , and it holds as a matter of law that all systems of type  $T$  have  $M$  at a time just in case they have  $P$  at the time.

We may call this “the Structure-Restricted Correlation Thesis” (or “the Restricted Correlation Thesis” for short).

It may have been noticed that neither this nor the correlation thesis speaks of “realization”.<sup>12</sup> The talk of “realization” is not metaphysically

---

<sup>11</sup> Thus, Post says, “Functional and intentional states are defined without regard to their physical or other realizations”, *The Faces of Existence*, p. 161. Also compare the earlier quotation from Block.

<sup>12</sup> As far as I know, the term “realization” was first used in something like its present sense by Hilary Putnam in “Minds and Machines”, in Sydney Hook, ed., *Dimensions of Mind* (New York: New York University Press, 1960).

neutral: the idea that mental properties are “realized” or “implemented” by physical properties carries with it a certain ontological picture of mental properties as derivative and dependent. There is the suggestion that when we look at concrete reality there is nothing over and beyond instantiations of physical properties and relations, and that the instantiation on a given occasion of an appropriate physical property in the right contextual (often causal) setting simply *counts as*, or *constitutes*, an instantiation of a mental property on that occasion. An idea like this is evident in the functionalist conception of a mental property as *extrinsically* characterized in terms of its “causal role”, where what fills this role is a physical (or, at any rate, nonmental) property (the latter property will then be said to “realize” the mental property in question). The same idea can be seen in the related functionalist proposal to construe a mental property as a “second-order property” consisting in the having of a physical property satisfying certain extrinsic specifications. We will recur to this topic later; however, we should note that someone who accepts either of the two correlation theses need not espouse the “realization” idiom. That is, it is *prima facie* a coherent position to think of mental properties as “first-order properties” in their own right, characterized by their intrinsic natures (e.g., phenomenal feel), which, as it happens, turn out to have nomological correlates in neural properties. (In fact, anyone interested in defending a serious dualist position on the mental should eschew the realization talk altogether and consider mental properties as first-order properties on a par with physical properties.) The main point of MR that is relevant to the antireductionist argument it has generated is just this: *mental properties do not have nomically coextensive physical properties, when the latter are appropriately individuated*. It may be that properties that are candidates for reduction must be thought of as being realized, or implemented, by properties in the prospective reduction base;<sup>13</sup> that is, if we think of certain properties as having their own intrinsic characterizations that are entirely independent of another set of properties, there is no hope of *reducing* the former to the latter. But this point needs to be argued, and will, in any case, not play a role in what follows.

Assume that property *M* is realized by property *P*. How are *M* and *P* related to each other and, in particular, how do they covary with each other? LePore and Loewer say this:<sup>14</sup>

---

<sup>13</sup> On this point see Robert Van Gulick, “Nonreductive Materialism and Intertheoretic Constraints”, in *Emergence or Reduction?*, ed. Ansgar Beckermann, Hans Flohr, and Jaegwon Kim (forthcoming from De Gruyter).

<sup>14</sup> “More on Making Mind Matter”, p. 179.

"The usual conception is that  $e$ 's being  $P$  realizes  $e$ 's being  $F$  iff  $e$  is  $P$  and there is a strong connection of some sort between  $P$  and  $F$ . We propose to understand this connection as a necessary connection which is *explanatory*. The existence of an explanatory connection between two properties is stronger than the claim that  $P \rightarrow F$  is physically necessary since not every physically necessary connection is explanatory."

Thus, LePore and Loewer require only that the realization base of  $M$  be *sufficient* for  $M$ , not both necessary and sufficient. This presumably is in response to MR: if pain is multiply realized in three ways as above, each of  $N_h$ ,  $N_r$ , and  $N_m$  will be sufficient for pain, and none necessary for it. This I believe is not a correct response, however; the correct response is not to weaken the joint necessity and sufficiency of the physical base, but rather to *relativize* it, as in the Restricted Correlation Thesis, with respect to species or structure types. For suppose we are designing a physical system that will instantiate a certain psychology, and let  $M_1, \dots, M_n$  be the psychological properties required by this psychology. The design process must involve the specification of an  $n$ -tuple of physical properties,  $P_1, \dots, P_n$ , all of them instantiable by the system, such that for each  $i$ ,  $P_i$  constitutes a *necessary and sufficient* condition in this system (and others of relevantly similar physical structure), not merely a sufficient one, for the occurrence of  $M_i$ . (Each such  $n$ -tuple of physical properties can be called a "physical realization" of the psychology in question.<sup>15</sup>) That is, for each psychological state we must design into the system a nomologically coextensive physical state. We must do this *if we are to control both the occurrence and non-occurrence of the psychological states involved*, and control of this kind necessary if we are to ensure that the physical device will properly instantiate the psychology. (This is especially clear if we think of building a computer; computer analogies loom large in our thoughts about "realization".)

But isn't it possible for multiple realization to occur "locally" as well? That is, we may want to avail ourselves of the flexibility of allowing a psychological state, or function, to be instantiated by alternative mechanisms within a single system. This means that  $P_i$  can be a *disjunction* of physical properties; thus,  $M_i$  is instantiated in the system in question at a time if and only if at least one of the disjuncts of  $P_i$  is instantiated at that time. The upshot of all this is that LePore and Loewer's condition that  $P \rightarrow M$  holds as a matter of law needs to be upgraded to the condition that, *relative to the*

<sup>15</sup> Cf. Hartry Field, "Mental Representation", in Block, *Readings in Philosophy of Psychology* (Cambridge: Harvard University Press, 1981), vol. 2.



*species or structure-type in question (and allowing  $P$  to be disjunctive),  $P \leftrightarrow M$  holds as a matter of law.*<sup>16</sup>

For simplicity let us suppose that pain is realized in three ways as above, by  $N_h$  in humans,  $N_r$  in reptiles, and  $N_m$  in Martians. The finitude assumption is not essential to any of my arguments: if the list is not finite, we will have an infinite disjunction rather than a finite one (alternatively, we can talk in terms of “sets” of such properties instead of their disjunctions). If the list is “open-ended”, that’s all right, too; it will not affect the metaphysics of the situation. We allowed above the possibility of a realization base of a psychological property itself being disjunctive; to get the discussion going, though, we will assume that these  $N$ s, the three imagined physical realization bases of pain, are not themselves disjunctive—or, at any rate, that their status as properties is not in dispute. The propriety and significance of “disjunctive properties” is precisely one of the principal issues we will be dealing with below, and it will make little difference just at what stage this issue is faced.

### III. Disjunctive Properties and Fodor’s Argument

An obvious initial response to the MR-based argument against reducibility is “the disjunction move”: Why not take the disjunction,  $N_h \vee N_r \vee N_m$ , as the single physical substrate of pain? In his 1967 paper, Putnam considers such a move but dismisses it out of hand: “Granted, in such a case the brain-state theorist can save himself by ad hoc assumptions (e.g., defining the disjunction of two states to be a single ‘physical-chemical state’), but this does not have to be taken seriously”.<sup>17</sup> Putnam gives no hint as to why he thinks the disjunction strategy does not merit serious consideration.

If there is something deeply wrong with disjunctions of the sort involved here, that surely isn’t obvious; we need to go beyond a sense of unease with such disjunctions and develop an intelligible rationale for banning them. Here is where Fodor steps in, for he appears to have an argument for disallowing disjunctions. As I see it, Fodor’s argument in “Special Sciences” depends crucially on the following two assumptions:

- (1) To reduce a special-science theory  $T_M$  to physical theory  $T_P$ , each “kind” in  $T_M$  (presumably, represented by a basic predicate of  $T_M$ ) must have a nomologically coextensive “kind” in  $T_P$ ;

---

<sup>16</sup> What of LePore and Loewer’s condition (ii), the requirement that the realization basis “explain” the realized property? Something like this explanatory relation may well be entailed by the realization relation; however, I do not believe it should be part of the definition of “realization”; that such an explanatory relation holds should be a consequence of the realization relation, not constitutive of it.

<sup>17</sup> “The Nature of Mental States”, p. 228 (in the Block volume).

(2) A disjunction of heterogeneous kinds is not itself a kind.

Point (1) is apparently prompted by the derivational model of intertheoretic reduction due to Ernest Nagel:<sup>18</sup> the reduction of  $T_2$  to  $T_1$  consists in the derivation of laws of  $T_2$  from the laws of  $T_1$ , in conjunction with “bridge” laws or principles connecting  $T_2$ -terms with  $T_1$ -terms. Although this characterization does not in general require that each  $T_2$ -term be correlated with a *coextensive*  $T_1$ -term, the natural thought is that the existence of  $T_1$ -coextensions for  $T_2$ -terms would in effect give us definitions of  $T_2$ -terms in  $T_1$ -terms, enabling us to rewrite  $T_2$ -laws exclusively in the vocabulary of  $T_1$ ; we could then derive these rewrites of  $T_2$ -laws from the laws of  $T_1$  (if they cannot be so derived, we can add them as additional  $T_1$ -laws—assuming both theories to be true).

Another thought that again leads us to look for  $T_1$ -coextensions for  $T_2$ -terms is this: for genuine reduction, the bridge laws must be construed as *property identities*, not mere *property correlations*—namely, we must be in a position to identify the property expressed by a given  $T_2$ -term (say, water-solubility) with a property expressed by a term in the reduction base (say, having a certain molecular structure). This of course requires that each  $T_2$ -term have a nomic (or otherwise suitably modalized) coextension in the vocabulary of the reduction base. To put it another way, ontologically significant reduction requires the reduction of higher-level *properties*, and this in turn requires (unless one takes an eliminativist stance) that they be identified with complexes of lower-level properties. Identity of properties of course requires, at a minimum, an appropriately modalized coextensivity.<sup>19</sup>

So assume  $M$  is a psychological kind, and let us agree that to reduce  $M$ , or to reduce the psychological theory containing  $M$ , we need a physical coextension,  $P$ , for  $M$ . But why should we suppose that  $P$  must be a physical “kind”? But what is a “kind”, anyway? Fodor explains this notion in terms of *law*, saying that a given predicate  $P$  is a “kind predicate” of a science just in case the science contains a law with  $P$  as its antecedent or consequent.<sup>20</sup> There are various problems with Fodor’s characterization, but we don’t need

---

<sup>18</sup> *The Structure of Science* (New York: Harcourt, Brace & World, 1961), chap. 11.

<sup>19</sup> My remarks here and the preceding paragraph assume that the higher-level theory requires no “correction” in relation to the base theory. With appropriate caveats and qualifications, they should apply to models of reduction that allow such corrections, or models that only require the deduction of a suitable analogue, or “image”, in the reduction base—as long as the departures are not so extreme as to warrant talk of replacement or elimination rather than reduction. Cf. Patricia Churchland, *Neurophilosophy* (Cambridge: The MIT Press, 1986), chap. 7.

<sup>20</sup> See “Special Sciences”, pp. 132–33 (in *Representations*).

to take its exact wording seriously; the main idea is that kinds, or kind predicates, of a science are those that figure in the laws of that science.

To return to our question, why should “bridge laws” connect kinds to kinds, in this special sense of “kind”? To say that bridge laws are “laws” and that, by definition, only kind predicates can occur in laws is not much of an answer. For that only invites the further question why “bridge laws” ought to be “laws”—what would be lacking in a reductive derivation if bridge laws were replaced by “bridge principles” which do not necessarily connect kinds to kinds.<sup>21</sup> But what of the consideration that these principles must represent property identities? Does this force on us the requirement that each reduced kind must find a coextensive kind in the reduction base? No; for it isn’t obvious why it isn’t perfectly proper to reduce kinds by identifying them with properties expressed by non-kind (disjunctive) predicates in the reduction base.

There is the following possible argument for insisting on kinds: if  $M$  is identified with non-kind  $Q$  (or  $M$  is reduced via a biconditional bridge principle “ $M \leftrightarrow Q$ ”, where  $Q$  is a non-kind),  $M$  could no longer figure in special science laws; e.g., the law, “ $M \rightarrow R$ ”, would in effect reduce to “ $Q \rightarrow R$ ”, and therefore loses its status as a law on account of containing  $Q$ , a non-kind.

I think this is a plausible response—at least, the beginning of one. As it stands, though, it smacks of circularity: “ $Q \rightarrow R$ ” is not a law because a non-kind,  $Q$ , occurs in it, and  $Q$  is a non-kind because it cannot occur in a law and “ $Q \rightarrow R$ ”, in particular, is not a law. What we need is an *independent* reason for the claim that the sort of  $Q$  we are dealing with under MR, namely a badly heterogeneous disjunction, is unsuited for laws.

This means that point (1) really reduces to point (2) above. For, given Fodor’s notion of a kind, (2) comes to this: disjunctions of heterogeneous kinds are unfit for laws. 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.<sup>22</sup> In the sections to follow, I hope to take some steps toward such a diagnosis and draw some implications which I believe are significant for the status of mentality.

---

<sup>21</sup> Fodor appears to assume that the requirement that bridge laws must connect “kinds” to “kinds” is part of the classic positivist conception of reduction. I don’t believe there is any warrant for this assumption, however.

<sup>22</sup> See Pereboom and Komblith, “The Metaphysics of Irreducibility” in which it is suggested that laws with disjunctive predicates are not “explanatory”. I think, though, that this suggestion is not fully developed there.

#### IV. Jade, Jadeite, and Nephrite

Let me begin with an analogy that will guide us in our thinking about multiply realizable kinds.

Consider *jade*: we are told that jade, as it turns out, is not a mineral kind, contrary to what was once believed; rather, jade is comprised of two distinct minerals with dissimilar molecular structures, *jadeite* and *nephrite*. Consider the following generalization:

(L) Jade is green

We may have thought, before the discovery of the dual nature of jade, that (L) was a law, a law about jade; and we may have thought, with reason, that (L) had been strongly confirmed by all the millions of jade samples that had been observed to be green (and none that had been observed not to be green). We now know better: (L) is really a conjunction of these two laws:

(L<sub>1</sub>) Jadeite is green

(L<sub>2</sub>) Nephrite is green

But (L) itself might still be a law as well; is that possible? It has the standard basic form of a law, and it apparently has the power to support counterfactuals: if anything were jade—that is, if anything were a sample of jadeite or of nephrite—then, in either case, it would follow, by law, that it was green. No problem here.

But there is another standard mark of lawlikeness that is often cited, and this is “projectibility”, the ability to be confirmed by observation of “positive instances”. Any generalized conditional of the form “All Fs are G” can be confirmed by the *exhaustion* of the class of Fs—that is, by eliminating all of its potential falsifiers. It is in this sense that we can verify such generalizations as “All the coins in my pockets are copper” and “Everyone in this room is either first-born or an only child”. Lawlike generalizations, however, are thought to have the following further property: observation of positive instances, Fs that are Gs, can strengthen our credence in the next F’s being G. It is this kind of instance-to-instance accretion of confirmation that is supposed to be the hallmark of lawlikeness; it is what explains the possibility of confirming a generalization about an indefinitely large class of items on the basis of a finite number of favorable observations. This rough characterization of projectibility should suffice for our purposes.

Does (L), “Jade is green”, pass the projectibility test? Here we seem to have a problem.<sup>23</sup> For we can imagine this: on re-examining the records of past observations, we find, to our dismay, that all the positive instances of (L), that is, all the millions of observed samples of green jade, turn out to have been samples of jadeite, and none of nephrite! If this should happen, we clearly would not, and should not, continue to think of (L) as well confirmed. All we have is evidence strongly confirming (L<sub>1</sub>), and none having anything to do with (L<sub>2</sub>). (L) is merely a conjunction of two laws, one well confirmed and the other with its epistemic status wholly up in the air. But all the millions of green jadeite samples *are* positive instances of (L): they satisfy both the antecedent and the consequent of (L). As we have just seen, however, (L) is not confirmed by them, at least not in the standard way we expect. And the reason, I suggest, is that jade is a true disjunctive kind, a disjunction of two heterogeneous nomic kinds which, however, is not itself a nomic kind.<sup>24</sup>

That disjunction is implicated in this failure of projectibility can be seen in the following way: inductive projection of generalizations like (L) with disjunctive antecedents would sanction a cheap, and illegitimate, confirmation procedure. For assume that “All *F*s are *G*” is a law that has been confirmed by the observation of appropriately numerous positive instances, things that are both *F* and *G*. But these are also positive instances of the generalization “All things that are *F* or *H* are *G*”, for any *H* you please. So, if you in general permit projection of generalizations with a disjunctive antecedent, this latter generalization is also well confirmed. But “All things that are *F* or *H* are *G*” logically implies “All *H*s are *G*”. Any statement implied by a well confirmed statement must itself be well confirmed.<sup>25</sup> So “All *H*s are *G*” is well confirmed—in fact, it is confirmed by the observation of *F*s that are *G*s!

One might protest: “Look, the very same strategy can be applied to something that is a genuine law. We can think of any nomic kind—say, being an emerald—as a disjunction, being an African emerald or a non-African emerald. This would make ‘All emeralds are green’ a conjunction of two

<sup>23</sup> The points to follow concerning disjunctive predicates were developed about a decade ago; however, I have just come across some related and, in some respects similar, points in David Owens’s interesting paper “Disjunctive Laws”, *Analysis* 49 (1989): 197–202. See also William Seager, “Disjunctive Laws and Supervenience”, *Analysis* 51 (1991): 93–98.

<sup>24</sup> This can be taken to define one useful sense of kind heterogeneity: two kinds are heterogeneous with respect to each other just in case their disjunction is not a kind.

<sup>25</sup> Note: this doesn’t say that for any *e*, if *e* is “positive evidence” for *h* and *h* logically implies *j*, then *e* is positive evidence for *j*. About the latter principle there is some dispute; see Carl G. Hempel, “Studies in the Logic of Confirmation”, reprinted in Hempel, *Aspects of Scientific Explanation* (New York: The Free Press, 1965), especially pp. 30–35; Rudolf Carnap, *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950), pp. 471–76.

laws, 'All African emeralds are green' and 'All non-African emeralds are green'. But surely this doesn't show there is anything wrong with the lawlikeness of 'All emeralds are green'. Our reply is obvious: the disjunction, "being an African emerald or non-African emerald", does not denote some heterogeneously disjunctive, nonnomic kind; it denotes a perfectly well-behaved nomic kind, that of being an emerald! There is nothing wrong with disjunctive predicates as such; the trouble arises when the kinds denoted by the disjoined predicates are heterogeneous, "wildly disjunctive", so that instances falling under them do not show the kind of "similarity", or unity, that we expect of instances falling under a single kind.

The phenomenon under discussion, therefore, is related to the simple maxim sometimes claimed to underlie inductive inference: "similar things behave in similar ways", "same cause, same effect", and so on. The source of the trouble we saw with instantial confirmation of "All jade is green" is the fact, or belief, that samples of jadeite and sample of nephrite do not exhibit an appropriate "similarity" with respect to each other to warrant inductive projections from the observed samples of jadeite to unobserved samples of nephrite. But similarity of the required sort presumably holds for African emeralds and non-African emeralds—at least, that is what we believe, and that is what makes the "disjunctive kind", being an African emerald or a non-African emerald, a single nomic kind. More generally, the phenomenon is related to the point often made about disjunctive properties: disjunctive properties, unlike conjunctive properties, do not guarantee similarity for instances falling under them. And similarity, it is said, is the core of our idea of a property. If that is your idea of a property, you will believe that there are no such things as disjunctive properties (or "negative properties"). More precisely, though, we should remember that properties are not inherently disjunctive or conjunctive any more than classes are inherently unions or intersections, and that any property can be expressed by a disjunctive predicate. Properties of course can be conjunctions, or disjunctions, of other properties. The point about disjunctive properties is best put as a closure condition on properties: the class of properties is not closed under disjunction (presumably, nor under negation). Thus, there may well be properties  $P$  and  $Q$  such that  $P$  or  $Q$  is also a property, but its being so doesn't follow from the mere fact that  $P$  and  $Q$  are properties.<sup>26</sup>

---

<sup>26</sup> On issues concerning properties, kinds, similarity, and lawlikeness, see W.V. Quine, "Natural Kinds" in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969); David Lewis, "New Work for a Theory of Universals", *Australasian Journal of Philosophy* 61 (1983): 347–77; D. M. Armstrong, *Universals* (Boulder, Colorado: Westview Press, 1989).

## V. Jade and Pain

Let us now return to pain and its multiple realization bases,  $N_h$ ,  $N_r$ , and  $N_m$ . I believe the situation here is instructively parallel to the case of jade in relation to jadeite and nephrite. It seems that we think of jadeite and nephrite as distinct kinds (and of jade not as a kind) because they are different chemical kinds. But why is their being distinct as chemical kinds relevant here? Because many important properties of minerals, we think, are supervenient on, and explainable in terms of, their microstructure, and chemical kinds constitute a microstructural taxonomy that is explanatorily rich and powerful. Microstructure is important, in short, because macrophysical properties of substances are determined by microstructure. These ideas make up our “metaphysics” of microdetermination for properties of minerals and other substances, a background of partly empirical and partly metaphysical assumptions that regulate our inductive and explanatory practices.

The parallel metaphysical underpinnings for pain, and other mental states in general, are, first, the belief, expressed by the Restricted Correlation Thesis, that pain, or any other mental state, occurs in a system when, and only when, appropriate physical conditions are present in the system, and, second, the corollary belief that significant properties of mental states, in particular nomic relationships amongst them, are due to, and explainable in terms of, the properties and causal-nomic connections among their physical “substrates”. I will call the conjunction of these two beliefs “the Physical Realization Thesis”.<sup>27</sup> Whether or not the micro-explanation of the sort indicated in the second half of the thesis amounts to a “reduction” is a question we will take up later. Apart from this question, though, the Physical Realization Thesis is widely accepted by philosophers who talk of “physical realization”, and this includes most functionalists; it is all but explicit in LePore and Loewer, for example, and in Fodor.<sup>28</sup>

Define a property,  $N$ , by disjoining  $N_h$ ,  $N_r$ , and  $N_m$ ; that is,  $N$  has a disjunctive definition,  $N_h \vee N_r \vee N_m$ . If we assume, with those who endorse the MR-based antireductionist argument, that  $N_h$ ,  $N_r$ , and  $N_m$  are a heterogeneous lot, we cannot make the heterogeneity go away merely by introducing a simpler expression, “ $N$ ”; if there is a problem with certain disjunctive

---

<sup>27</sup> This term is a little misleading since the two subtheses have been stated without the term “realization” and may be acceptable to those who would reject the “realization” idiom in connection with the mental. I use the term since we are chiefly addressing philosophers (mainly functionalists) who construe the psychophysical relation in terms of realization, rather than, say, emergence or brute correlation.

<sup>28</sup> See “Special Sciences”, and “Making Mind Matter More”, *Philosophical Topics* 17 (1989): 59–79.

properties, it is not a *linguistic* problem about the form of expressions used to refer to them.

Now, we put the following question to Fodor and like-minded philosophers: If pain is nomically equivalent to  $N$ , the property claimed to be wildly disjunctive and obviously nonnommic, *why isn't pain itself equally heterogeneous and nonnommic as a kind?* Why isn't pain's relationship to its realization bases,  $N_h$ ,  $N_r$ , and  $N_m$  analogous to jade's relationship to jadeite and nephrite? If jade turns out to be nonnommic on account of its dual "realizations" in distinct microstructures, why doesn't the same fate befall pain? After all, the group of actual and nomologically possible realizations of pain, as they are described by the MR enthusiasts with such imagination, is far more motley than the two chemical kinds comprising jade.

I believe we should insist on answers to these questions from those functionalists who view mental properties as "second-order" properties, i.e., properties that consist in having a property with a certain functional specification.<sup>29</sup> Thus, pain is said to be a second-order property in that it is the *property of having some property with a certain specification* in terms of its typical causes and effects and its relation to other mental properties; call this "specification  $H$ ". The point of MR, on this view, is that there is more than one property that meets specification  $H$ —in fact, an open-ended set of such properties, it will be said. But pain itself, it is argued, is a more abstract but well-behaved property at a higher level, namely the property of having one of these properties meeting specification  $H$ . It should be clear why a position like this is vulnerable to the questions that have been raised. For the property of having property  $P$  is exactly identical with  $P$ , and the property of having *one* of the properties,  $P_1, P_2, \dots, P_n$  is exactly identical with the disjunctive property,  $P_1 \vee P_2 \vee \dots \vee P_n$ . On the assumption that  $N_h, N_r$ , and  $N_m$  are all the properties satisfying specification  $H$ , the property of having a property with  $H$ , namely pain, is none other than the property of having either  $N_h$  or  $N_r$  or  $N_m$ <sup>30</sup>—namely, the *disjunctive* property,  $N_h \vee N_r \vee N_m$ ! We cannot hide the disjunctive character of pain behind the second-order *expression*, "the property of having a property with specification  $H$ ". 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. It is difficult to see how one could have it both ways—that is, to castigate  $N_h \vee N_r \vee N_m$  as unacceptably disjunctive while insisting on the integrity of pain as a scientific kind.

Moreover, when we think about making projections over pain, very much the same worry should arise about their propriety as did for jade. Consider a

<sup>29</sup> See, e.g., Block, "Can the Mind Change the World?", p. 155.

<sup>30</sup> We might keep in mind the close relationship between disjunction and the existential quantifier standardly noted in logic textbooks.



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.<sup>31</sup> “Pains cause anxiety reactions” may turn out to possess no more unity as a scientific law than does “Jade is green”.

Suppose that in spite of all this Fodor insists on defending pain as a nomic kind. It isn’t clear that that would be a viable strategy. For he would then owe us an explanation of why the “wildly disjunctive” *N*, which after all is equivalent to pain, is not a nomic kind. If a predicate is nomically equivalent to a well-behaved predicate, why isn’t that enough to show that it, too, is well behaved, and expresses a well-behaved property? To say, as Fodor does,<sup>32</sup> that “it is a law that...” is “intensional” and does not permit substitution of equivalent expressions (“equivalent” in various appropriate senses) is merely to locate a potential problem, not to resolve it.

Thus, the nomicity of pain may lead to the nomicity of *N*; but this isn’t very interesting. For given the Physical Realization Thesis, and the priority of the physical implicit in it, our earlier line of argument, leading from the nonnomicity of *N* to the nonnomicity of pain, is more compelling. We must, I think, take seriously the reasoning leading to the conclusion that pain, and other mental states, might turn out to be nonnomic. If this turns out to be the case, it puts in serious jeopardy Fodor’s contention that its physical irreducibility renders psychology an autonomous special science. If pain fails to be nomic, it is not the sort of property in terms of which laws can be formulated; and “pain” is not a predicate that can enter into a scientific theory that seeks to formulate causal laws and causal explanations. And the same goes for all multiply realizable psychological kinds—which, according to MR, means *all* psychological kinds. There are no scientific theories of jade, and we don’t need any; if you insist on having one, you can

---

<sup>31</sup> It may be a complicated affair to formulate this argument within certain functionalist schemes; if, for example, mental properties are functionally defined by Ramseyfying a total psychological theory, it will turn out that humans and Martians cannot share any psychological state unless the same total psychology (including the putative law in question) is true (or held to be true) for both.

<sup>32</sup> “Special Sciences”, p. 140 (in *Representations*).

help yourself with the *conjunction* of the theory of jadeite and the theory of nephrite. In the same way, there will be theories about human pains (instances of  $N_h$ ), reptilian pains (instances of  $N_r$ ), and so on; but there will be no unified, integrated theory encompassing all pains in all pain-capable organisms, only a conjunction of pain theories for appropriately individuated biological species and physical structure-types. Scientific psychology, like the theory of jade, gives way to a conjunction of structure-specific theories. If this is right, the correct conclusion to be drawn from the MR-inspired antireductionist argument is not the claim that psychology is an irreducible and autonomous science, but something that contradicts it, namely that it cannot be a science with a unified subject matter. This is the picture that is beginning to emerge from MR when combined with the Physical Realization Thesis.

These reflections have been prompted by the analogy with the case of jade; it is a strong and instructive analogy, I think, and suggests the possibility of a general argument. In the following section I will develop a direct argument, with explicit premises and assumptions.

## VI. Causal Powers and Mental Kinds

One crucial premise we need for a direct argument is a constraint on concept formation, or kind individuation, in science that has been around for many years; it has lately been resurrected by Fodor in connection with content externalism.<sup>33</sup> A precise statement of the constraint may be difficult and controversial, but its main idea can be put as follows:

[Principle of Causal Individuation of Kinds] Kinds in science are individuated on the basis of causal powers; that is, objects and events fall under a kind, or share in a property, insofar as they have similar causal powers.

I believe this is a plausible principle, and it is, in any case, widely accepted.

We can see that this principle enables us to give a specific interpretation to the claim that  $N_h$ ,  $N_r$ , and  $N_m$  are *heterogeneous* as kinds: the claim must mean that they are *heterogeneous as causal powers*—that is, they are diverse as causal powers and enter into diverse causal laws. This must mean, given the Physical Realization Thesis, that pain itself can show no more unity as a causal power than the disjunction,  $N_h \vee N_r \vee N_m$ . This becomes especially clear

---

<sup>33</sup> See, e.g., Carl G. Hempel, *Fundamentals of Concept Formation in Empirical Science* (Chicago: University of Chicago Press, 1952); W.Q. Quine, "Natural Kinds". Fodor gives it an explicit statement in *Psychosemantics* (Cambridge: MIT Press, 1988), chap. 2. A principle like this is often invoked in the current externalism/internalism debate about content; most principal participants in this debate seem to accept it.

if we set forth the following principle, which arguably is implied by the Physical Realization Thesis (but we need not make an issue of this here):

[The Causal Inheritance Principle] 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*.<sup>34</sup>

It is important to bear in mind that this principle only concerns the causal powers of *individual instances* of *M*; it does not identify the causal powers of mental property *M in general* with the causal powers of some physical property *P*; such identification is precluded by the multiple physical realizability of *M*.

Why should we accept this principle? Let us just note that to deny it would be to accept *emergent* causal powers: causal powers that magically emerge at a higher-level and of which there is no accounting in terms of lower-level properties and their causal powers and nomic connections. This leads to the notorious problem of "downward causation" and the attendant violation of the causal closure of the physical domain.<sup>35</sup> I believe that a serious physicalist would find these consequences intolerable.

It is clear that the Causal Inheritance Principle, in conjunction with the Physical Realization Thesis, has the consequence that mental kinds cannot satisfy the Causal Individuation Principle, and this effectively rules out mental kinds as scientific kinds. The reasoning is simple: instances of *M* that are realized by the same physical base must be grouped under one kind, since *ex hypothesi* the physical base is a causal kind; and instances of *M* with different realization bases must be grouped under distinct kinds, since, again *ex hypothesi*, these realization bases are distinct as causal kinds. Given that mental kinds are realized by diverse physical causal kinds, therefore, it follows that mental kinds are not causal kinds, and hence are disqualified as proper scientific kinds. Each mental kind is sundered into as many kinds as there are physical realization bases for it, and the psychology as a science with disciplinary unity turns out to be an impossible project.

---

<sup>34</sup> A principle like this is sometimes put in terms of "supervenience" and "supervenience base" rather than "realization" and "realization base". See my "Epiphenomenal and Supervenient Causation", *Midwest Studies in Philosophy* 9 (1984): 257–70. Fodor appears to accept just such a principle of supervenient causation for mental properties in chap. 2 of his *Psychosemantics*. In "The Metaphysics of Irreducibility" Pereboom and Kornblith appear to reject it.

<sup>35</sup> For more details see my "'Downward Causation' in Emergentism and Nonreductive Physicalism", forthcoming in *Emergence or Reduction?*, ed. Beckermann, Flohr, and Kim, and "The Nonreductivist's Troubles with Mental Causation", forthcoming in *Mental Causation*, ed. John Heil and Alfred Mele (Oxford University Press).

What is the relationship between this argument and the argument adumbrated in our reflections based on the jade analogy? At first blush, the two arguments might seem unrelated: the earlier argument depended chiefly on epistemological considerations, considerations on inductive projectibility of certain predicates, whereas the crucial premise of the second argument is the Causal Kind Individuation Principle, a broadly metaphysical and methodological principle about science. I think, though, that the two arguments are closely related, and the key to seeing the relationship is this: causal powers involve laws, and laws are regularities that are projectible. Thus, if pain (or jade) is not a kind over which inductive projections can be made, it cannot enter into laws, and therefore cannot qualify as a causal kind; and this disqualifies it as a scientific kind. If this is right, the jade-inspired reflections provide a possible rationale for the Causal Individuation Principle. Fleshing out this rough chain of reasoning in precise terms, however, goes beyond what I can attempt in this paper.

## VII. The Status of Psychology: Local Reductions

Our conclusion at this point, therefore, is this: If MR is true, psychological kinds are not scientific kinds. What does this imply about the status of psychology as a science? Do our considerations show that psychology is a pseudo-science like astrology and alchemy? Of course not. The crucial difference, from the metaphysical point of view, is that psychology has physical realizations, but alchemy does not. To have a physical realization is to be physically grounded and explainable in terms of the processes at an underlying level. In fact, if each of the psychological kinds posited in a psychological theory has a physical realization for a fixed species, the theory can be “locally reduced” to the physical theory of that species, in the following sense. Let  $S$  be the species involved; for each law  $L_m$  of psychological theory  $T_m$ ,  $S \rightarrow L_m$  (the proposition that  $L_m$  holds for members of  $S$ ) is the “ $S$ -restricted” version of  $L_m$ ; and  $S \rightarrow T_m$  is the  $S$ -restricted version of  $T_m$ , the set of all  $S$ -restricted laws of  $T_m$ . We can then say that  $T_m$  is “locally reduced” for species  $S$  to an underlying theory,  $T_p$ , just in case  $S \rightarrow T_m$  is reduced to  $T_p$ . And the latter obtains just in case each  $S$ -restricted law of  $T_m$ ,  $S \rightarrow L_m$ ,<sup>36</sup> is derivable from the laws of the reducing theory  $T_p$ , taken together with bridge laws. What bridge laws suffice to guarantee the derivation? Obviously, an array of  $S$ -restricted bridge laws of the form,  $S \rightarrow (M_i \leftrightarrow P_i)$ , for each mental kind  $M_i$ . Just as unrestricted psychophysical bridge laws can underwrite a “global” or “uniform” reduction of psychology, species- or structure-restricted bridge laws sanction its “local” reduction.

---

<sup>36</sup> Or an appropriately corrected version thereof (this qualification applies to the bridge laws as well).

If the same psychological theory is true of humans, reptiles, and Martians, the psychological kinds posited by that theory must have realizations in human, reptilian, and Martian physiologies. This implies that the theory is locally reducible in three ways, for humans, reptiles, and Martians. If the dependence of the mental on the physical means anything, it must mean that the regularities posited by this common psychology must have divergent physical explanations for the three species. The very idea of physical realization involves the possibility of physically explaining psychological properties and regularities, and the supposition of multiple such realizations, namely MR, involves a commitment to the possibility of multiple explanatory reductions of psychology.<sup>37</sup> The important moral of MR we need to keep in mind is this: *if psychological properties are multiply realized, so is psychology itself*. If physical realizations of psychological properties are a “wildly heterogeneous” and “unsystematic” lot, psychological theory itself must be realized by an equally heterogeneous and unsystematic lot of physical theories.

I am inclined to think that multiple local reductions, rather than global reductions, are the rule, even in areas in which we standardly suppose reductions are possible. I will now deal with a possible objection to the idea of local reduction, at least as it is applied to psychology. The objection goes like this: given what we know about the differences among members of a single species, even species are too wide to yield determinate realization bases for psychological states, and given what we know about the phenomena of maturation and development, brain injuries, and the like, the physical bases of mentality may change even for a single individual. This throws into serious doubt, continues the objection, the availability of species-restricted bridge laws needed for local reductions.

The point of this objection may well be correct as a matter of empirical fact. Two points can be made in reply, however. First, neurophysiological research goes on because there is a shared, and probably well grounded, belief among the workers that there are not huge individual differences within a species in the way psychological kinds are realized. Conspecifics must show important physical-physiological similarities, and there probably is good reason for thinking that they share physical realization bases to a sufficient degree to make search for species-wide neural substrates for mental states feasible and rewarding. Researchers in this area evidently aim for

---

<sup>37</sup> In “Special Sciences” and “Making Mind Matter More” Fodor appears to accept the local reducibility of psychology and other special sciences. But he uses the terminology of local *explanation*, rather than reduction, of psychological regularities in terms of underlying microstructure. I think this is because his preoccupation with Nagelian uniform reduction prevents him from seeing that this is a form of inter-theoretic reduction if anything is.

neurobiological explanations of psychological capacities and processes that are generalizable over all or most ("normal") members of a given species.

Second, even if there are huge individual differences among conspecifics as to how their psychology is realized, that does not touch the metaphysical point: as long as you believe in the Physical Realization Thesis, you must believe that every organism or system with mentality falls under a physical structure-type such that its mental states are realized by determinate physical states of organisms with that structure. It may be that these structures are so finely individuated and so few *actual* individuals fall under them that research into the neural bases of mental states in these structures is no longer worthwhile, theoretically or practically. What we need to recognize here is that the scientific possibility of, say, human psychology is a contingent fact (assuming it is a fact); it depends on the fortunate fact that individual humans do not show huge physiological-biological differences that are psychologically relevant. But if they did, that would not change the metaphysics of the situation one bit; it would remain true that the psychology of each of us was determined by, and locally reducible to, his neurobiology.

Realistically, there are going to be psychological differences among individual humans: it is a commonsense platitude that no two persons are exactly alike—either physically or psychologically. And individual differences may be manifested not only in particular psychological facts but in psychological regularities. If we believe in the Physical Realization Thesis, we must believe that our psychological differences are rooted in, and explainable by, our physical differences, just as we expect our psychological similarities to be so explainable. Humans probably are less alike among themselves than, say, tokens of a Chevrolet model.<sup>38</sup> And psychological laws for humans, at a certain level of specificity, must be expected to be statistical in character, not deterministic—or, if you prefer, "ceteris paribus laws" rather than "strict laws". But this is nothing peculiar to psychology; these remarks surely apply to human physiology and anatomy as much as human psychology. In any case, none of this affects the metaphysical point being argued here concerning microdetermination and microreductive explanation.

---

<sup>38</sup> Compare J. J. C. Smart's instructive analogy between biological organisms and super-heterodyne radios, in *Philosophy and Scientific Realism* (London: Routledge & Kegan Paul, 1963), pp. 56–57. Smart's conception of the relation between physics and the special sciences, such as biology and psychology, is similar in some respects to the position I am defending here.

### VIII. Metaphysical Implications

But does local reduction have any interesting philosophical significance, especially in regard to the status of mental properties? If a psychological property has been multiply locally reduced, does that mean that the property itself has been reduced? Ned Block has raised just such a point, arguing that species-restricted reductionism (or species-restricted type physicalism) “sidesteps the main metaphysical question: ‘What is common to the pains of dogs and people (and all other species) in virtue of which they are pains?’”<sup>39</sup>

Pereboom and Kornblith elaborate on Block’s point as follows:

“...even if there is a single type of physical state that normally realizes pain in each type of organism, or in each structure type, this does not show that pain, *as a type of mental state*, is reducible to physical states. Reduction, in the present debate, must be understood as reduction of types, since the primary object of reductive strategies is explanations and theories, and explanations and theories quantify over types....The suggestion that there are species-specific reductions of pain results in the claim that pains in different species have nothing in common. But this is just a form of eliminativism.”<sup>40</sup>

There are several related but separable issues raised here. But first we should ask: Must all pains have “something in common” in virtue of which they are pains?

According to the phenomenological conception of pain, all pains do have something in common: they all *hurt*. But as I take it, those who hold this view of pain would reject any reductionist program, independently of the issues presently on hand. Even if there were a species-invariant uniform bridge law correlating pains with a single physical substrate across all species and structures, they would claim that the correlation holds as a brute, unexplainable matter of fact, and that pain as a qualitative event, a “raw feel”, would remain irreducibly distinct from its neural substrate. Many emergentists apparently held a view of this kind.

---

<sup>39</sup> “Introduction: What is Functionalism?” in *Readings in Philosophy of Psychology*, pp. 178–79.

<sup>40</sup> In their “The Metaphysics of Irreducibility”. See also Ronald Endicott, “The Species-Specific Strategy”, forthcoming. In personal correspondence Earl Conee and Joe Mendola have raised similar points. There is a useful discussion of various metaphysical issues relating to MR in Cynthia Macdonald, *Mind-Body Identity Theories* (London and New York: Routledge, 1989).

I presume that Block, and Pereboom and Kornblith, are speaking not from a phenomenological viewpoint of this kind but from a broadly functionalist one. But from a functionalist perspective, it is by no means clear how we should understand the question “What do all pains have in common in virtue of which they are all pains?” Why should all pains have “something in common”? As I understand it, at the core of the functionalist program is the attempt to explain the meanings of mental terms *relationally*, in terms of inputs, outputs, and connections with other mental states. And on the view, discussed briefly earlier, that mental properties are second-order properties, pain is the property of having a property with a certain functional specification  $H$  (in terms of inputs, outputs, etc.). This yields a short answer to Block’s question: what all pains have in common is the pattern of connections as specified by  $H$ . The local reductionist is entitled to that answer as much as the functionalist is. Compare two pains, an instance of  $N_h$  and one of  $N_m$ : what they have in common is that each is an instance of a property that realizes pain—that is, they exhibit the same pattern of input-output-other internal state connections, namely the pattern specified by  $H$ .

But some will say: “But  $H$  is only an *extrinsic* characterization; what do these instances of pain have in common that is *intrinsic* to them?” The local reductionist must grant that on his view there is nothing intrinsic that all pains have in common in virtue of which they are pains (assuming that  $N_h$ ,  $N_r$ , and  $N_m$  “have nothing intrinsic in common”). But that is also precisely the consequence of the functionalist view. That, one might say, is the whole point of functionalism: the functionalist, especially one who believes in MR, would not, and should not, look for something common to all pains over and above  $H$  (the heart of functionalism, one might say, is the belief that mental states have no “intrinsic essence”).

But there is a further question raised by Block et al.: What happens to properties that have been locally reduced? Are they still with us, distinct and separate from the underlying physical-biological properties? Granted: human pain is reduced to  $N_h$ , Martian pain to  $N_m$ , and so forth, but what of *pain itself*? It remains unreduced. Are we still stuck with the dualism of mental and physical properties?

I will sketch two possible ways of meeting this challenge. First, recall my earlier remarks about the functionalist conception of mental properties as second-order properties: pain is *the property of having a property with specification  $H$* , and, given that  $N_h$ ,  $N_r$ , and  $N_m$  are the properties meeting  $H$ , pain turns to be the disjunctive property,  $N_h \vee N_r \vee N_m$ . If you hold the second-order property view of mental properties, pain has been reduced to, and survives as, this disjunctive physical kind. Quite apart from considerations of local reduction, the very conception of pain you hold commits you to the conclusion that pain is a disjunctive kind, and if you accept any form of re-



spectable physicalism (in particular, the Physical Realization Thesis), it is a disjunctive *physical* kind. And even if you don't accept the view of mental properties as second-order properties, as long as you are comfortable with disjunctive kinds and properties, you can, in the aftermath of local reduction, identify pain with the disjunction of its realization bases. On this approach, then, you have another, more direct, answer to Block's question: what all pains have in common is that they all fall under the disjunctive kind,  $N_h \vee N_r \vee N_m$ .

If you are averse to disjunctive kinds, there is another more radical, and in some ways more satisfying, approach. The starting point of this approach is the frank acknowledgement that MR leads to the conclusion that pain as a property or kind must go. Local reduction after all is reduction, and to be reduced is to be eliminated as an *independent* entity. You might say: global reduction is different in that it is also *conservative*—if pain is globally reduced to physical property  $P$ , pain survives as  $P$ . But it is also true that under local reduction, pain survives as  $N_h$  in humans, as  $N_r$  in reptiles, and so on. It must be admitted, however, that pain as a kind does not survive multiple local reduction. But is this so bad?

Let us return to jade once again. Is jade a *kind*? We know it is not a mineral kind; but is it any kind of a kind? That of course depends on what we mean by "kind". There are certain shared criteria, largely based on observable macroproperties of mineral samples (e.g., hardness, color, etc.), that determine whether something is a sample of jade, or whether the predicate "is jade" is correctly applicable to it. What all samples of jade have in common is just these observable macrophysical properties that define the applicability of the predicate "is jade". In this sense, speakers of English who have "jade" in their repertoire associate the same *concept* with "jade"; and we can recognize the existence of the concept of jade and at the same time acknowledge that the concept does not pick out, or answer to, a property or kind in the natural world.

I think we can say something similar about pain and "pain": there are shared criteria for the application of the predicate "pain" or "is in pain", and these criteria may well be for the most part functionalist ones. These criteria generate for us a *concept of pain*, a concept whose clarity and determinacy depend, we may assume, on certain characteristics (such as explicitness, coherence, and completeness) of the criteria governing the application of "pain". But the concept of pain, on this construal, need not pick out an objective kind any more than the concept of jade does.

All this presupposes a distinction between concepts and properties (or kinds). Do we have such a distinction? I believe we do. Roughly, concepts are in the same ball park as predicates, meanings (perhaps, something like Fregean *Sinnen*), ideas, and the like; Putnam has suggested that concepts be

identified with “synonymy classes of predicates”,<sup>41</sup> and that comes close enough to what I have in mind. Properties and relations, on the other hand, are “out there in the world”; they are features and characteristics of things and events in the world. They include fundamental physical magnitudes and quantities, like mass, energy, size, and shape, and are part of the causal structure of the world. The property of being water is arguably identical with the property of being H<sub>2</sub>O, but evidently the concept of water is distinct from the concept of H<sub>2</sub>O (Socrates had the former but not the latter). Most of us would agree that ethical predicates are meaningful, and that we have the concepts of “good”, “right”, etc.; however, it is a debatable issue, and has lately been much debated, whether there are such properties as goodness and rightness.<sup>42</sup> If you find that most of these remarks make sense, you understand the concept-property distinction that I have in mind. Admittedly, this is all a little vague and programmatic, and we clearly need a better articulated theory of properties and concepts; but the distinction is there, supported by an impressively systematic set of intuitions and philosophical requirements.<sup>43</sup>

But is this second approach a form of mental eliminativism? In a sense it is: as I said, on this approach no properties in the world answer to general, species-unrestricted mental concepts. But remember: there still are pains, and we sometimes are in pain, just as there still are samples of jade. We must also keep in mind that the present approach is not, in its ontological implications, a form of the standard mental eliminativism currently on the scene.<sup>44</sup> Without elaborating on what the differences are, let us just note a few important points. First, the present view does not take away species-restricted mental properties, e.g., human pain, Martian pain, canine pain, and the rest, although it takes away “pain as such”. Second, while the standard eliminativism consigns mentality to the same ontological limbo to which phlogiston, witches, and magnetic effluvia, have been dispatched, the position I have been sketching views it on a par with jade, tables, and adding machines. To see jade as a nonkind is not to question the existence of jade, or the legitimacy and utility of the concept of jade. Tables do not constitute a scientific kind; there are no laws about tables as such, and being a table is not a causal-explanatory kind. But that must be sharply distinguished from the false claim that there are no tables. The same goes for pains. These points suggest

---

<sup>41</sup> In “The Nature of Mental States”.

<sup>42</sup> I of course have in mind the controversy concerning moral realism; see essays in Geoffrey Sayre-McCord, ed., *Essays on Moral Realism* (Ithaca: Cornell University Press, 1988).

<sup>43</sup> On concepts and properties, see, e.g., Hilary Putnam, “On Properties”, *Mathematics, Matter and Method* (Cambridge: Cambridge University Press, 1975); Mark Wilson, “Predicate Meets Property”, *Philosophical Review* 91 (1982): 549–90, especially, section III.

<sup>44</sup> Such as the versions favored by W.V. Quine, Stephen Stich, and Paul Churchland.

the following difference in regard to the status of psychology: the present view allows, and in fact encourages, “species-specific psychologies”, but the standard eliminativism would do away with all things psychological—species-specific psychologies as well as global psychology.<sup>45</sup>

To summarize, then, the two metaphysical schemes I have sketched offer these choices: either we allow disjunctive kinds and construe pain and other mental properties as such kinds, or else we must acknowledge that our general mental terms and concepts do not pick out properties and kinds in the world (we may call this “mental property irrealism”). I should add that I am not interested in promoting either disjunctive kinds or mental irrealism, a troubling set of choices to most of us. Rather, my main interest has been to follow out the consequences of MR and try to come to terms with them within a reasonable metaphysical scheme.

I have already commented on the status of psychology as a science under MR. As I argued, MR seriously compromises the disciplinary unity and autonomy of psychology as a science. But that does not have to be taken as a negative message. In particular, the claim does not imply that a scientific study of psychological phenomena is not possible or useful; on the contrary, MR says that psychological processes have a foundation in the biological and physical processes and regularities, and it opens the possibility of enlightening explanations of psychological processes at a more basic level. It is only that at a deeper level, psychology becomes sundered by being multiply locally reduced. However, species-specific psychologies, e.g., human psychology, Martian psychology, etc., can all flourish as scientific theories. Psychology remains *scientific*, though perhaps not *a science*. If you insist on having a global psychology valid for all species and structures, you can help yourself with that, too; but you must think of it as a *conjunction* of species-restricted psychologies and be careful, above all, with your inductions.<sup>46</sup>

---

<sup>45</sup> The approach to the mind-body problem being adumbrated here is elaborated in my “Functionalism as Mental Irrealism” (in preparation).

<sup>46</sup> This paper is descended from an unpublished paper, “The Disunity of Psychology as a Working Hypothesis?”, which was circulated in the early 1980s. I am indebted to the following persons, among others, for helpful comments: Fred Feldman, Hilary Kornblith, Barry Loewer, Brian McLaughlin, Joe Mendola, Marcelo Sabates, and James Van Cleve.