

Scattered Remarks On Multiple Realizability[§]

Hong Yu Wong

1. Ontological Candor

Since his ontological conversion on the road to supervenience physicalism, John Heil has been proselytizing about ontological candor, and peddling his and Charlie Martin's dual-aspect theory of properties as a somewhat magical cure for "the current malaise in the philosophy of mind" (Heil 1999, 2000). Properties have both dispositional *and* qualitative aspects, and these aspects only come apart in *thought*; in other words, they only come apart in analysis. Heil and Martin will probably balk at my characterization of their theory of properties as "dual-aspect"; they dislike talk in terms of "aspects", for then the qualitative and dispositional "aspects" seem like independent features of properties. Note that on the Heil-Martin theory, properties look just as they would on the pure dispositions account when viewed (or analyzed) from a "dispositional" angle. Heil and Martin tell us no story about why or how properties have both dispositional *and* qualitative aspects, only that properties *have* both dispositional and qualitative aspects. It is a stipulation that comes across as somewhat arbitrary, since the only reason to embrace the dual-aspect theory is its allowing for a physicalistic account of consciousness by fiat.¹ Heil and Martin want the dispositional account, where properties are exhausted by their causal role, and they also want to have properties with intrinsic qualities. They want to diffuse this tension between the dispositional and the qualitative by stipulating that they are but two sides of the same coin. Heil complains of "top-down ontologizing" that his colleagues delight in, yet fails to notice that he is doing exactly the same thing. He is mistaken in thinking that somehow his dual-aspect theory of properties is more adequate than other theories of properties *simpliciter*. The ontology cannot be independently justified; the ends justify the means. But his point about ontological candor is a point well taken.²

Take the multiple realizability intuition for example. There are many questions that one is confronted with once one begins filling in the ontological details. A common construal of mental properties on the functionalist account is as second-order functional properties. On the standard nonreductive physicalist (NRP) account, these second-order functional properties are not epiphenomenal, i.e. they are causally efficacious, but from wherein do they derive their causal powers? If they inherit their causal powers from their first-order realizers why not just reductively identify mental properties with their first-order realizers? The proponent of multiple realizability answers that type identity requires nomological biconditionals³ which are not available, because mental properties are

[§] Submitted for the CUNY 5th Annual Graduate Student Philosophy Conference, 2001.

¹ I am perhaps being unfair to Heil and Martin here. Unfortunately, the only example Heil applies the dual-aspect theory to in his published corpus is that of consciousness. Perhaps Heil can say in defense that because scientific kinds are generally causal kinds, science has tended to characterize physical properties functionally, and hence we have no good examples of qualitative aspects of physical properties.

² On ontological candor, see Heil and Martin (1998), Robb and Heil (1998) and Stewart (1997).

³ It is unclear that nomological biconditionals are necessary for type reduction unless one accepts the classical Nagel model of reduction. Jaegwon Kim has argued that the Nagel model is neither necessary nor

realized by multiple distinct first-order physical realizers. Each distinct physical realizer is sufficient for the instantiation of the mental property, but none are necessary. Now, depending on one's theory of properties, various difficulties emerge for this account of mental properties. If we take properties to be individuated by their causal contribution (causal powers),⁴ then an apparent contradiction ensues. Imagine a mental property M , say pain (to take a passé example), which is realized distinctly in human, crocodile and octopus neural circuits, N_h , N_c , and N_o respectively. If indeed the first-order physical realizers are distinct, it must be because each physical realizer contributes different causal powers, since we assume properties to be exhausted by their causal roles. So $N_h \neq N_c \neq N_o$. But $N_h = M$, $N_c = M$, and $N_o = M$, so by transitivity of "=", $N_h = N_c = N_o$. Apparent contradictions like this evaporate once one adopts a broadly Humean view of causation.

Ontology is the most general architectonic within which we make sense of human activity, inquiry and knowledge. It is how we carve up nature and plumb the fundamental categories. There is a certain relativity to ontological schemes, and different schemes buy you different things. I do not think ontological schemes can be independently justified. Our task is to try out alternative ontologies and find the ontological scheme of best fit. In this task we are constrained by both our pretheoretic intuitions about the world, as well as current scientific knowledge about the world.

In what follows, I will reexamine the centerpiece of nonreductive physicalism—the multiple realizability (MR) argument to irreducibility—and reevaluate its viability *qua* ontology. As with all classic philosophical arguments and examples, MR is given a unique twist by every philosopher who wields it. Some, like Fodor, claim that it is an argument for token physicalism and blocks type reductionism; others, like Kim, have argued that because scientific kinds are individuated by the causal roles they play, and since multiply realized kinds are individuated in the exact same manner, type reduction follows.⁵ To the unsuspecting reader (say a student philosopher), this seems patently contradictory. With a quizzical look she asks, “so multiple realizability leads to both token and type reduction? No?”⁶ Often implicit in discussions of the tenability of NRP and the consequences of MR are the favorite ontologies of individual philosophers (this

sufficient for reduction (Kim, 1998, chapter 4). In particular, the existence of Nagelian bridge laws connecting a reducing and a reduced theory is consistent with the doctrine of the British emergentists, who allowed for nonexplanatory, brute bridge laws which were to be accepted with “natural piety”. If only nonexplanatory bridge laws link a reduced and a reducing theory, then the reduced theory has not really been reduced, since the reduced theory cannot be reductively explained in terms of the reducing theory.

⁴ See Shoemaker (1980).

⁵ This is the gist of the functional model of reduction that Kim currently espouses. Jaegwon Kim began as one of the foremost proponents of supervenience physicalism, the doctrine that mental supervenes on the physical but is *irreducible* to the physical. He has since moved away from nonreductive physicalism toward a robust type-reductionism based on a functional model of reduction (Kim 1998, 1999). See Horgan (1996) for a review of Kim's positions on the mind-body problem through Kim (1993) and essays in the section “Mental Causation, Reduction and Supervenience” (pp. 83-208) in *Philosophical Perspectives*, Vol. 11 (1997) for reviews of Kim's particular brand of functionalism.

⁶ The resolution of this issue hinges on one's exact ontology of events. I will return to consider differences between Davidsonian and Kimian events later in this paper, and the relation between their ontology of events and their solution to the mind-body problem.

gets us back to the ontological candor point). Unfortunately (or fortunately⁷), the majority of present day philosophers are closet metaphysicians who maintain a *hush-hush* attitude toward ontological issues. I will attempt to tease out these secret ontologies and show just how crucial these unspoken (and unwritten) ontologies are to each perspective on MR and NRP (and of course to any question in the philosophy of mind). In fact, I like to think that individual ontologies drive individual philosophical enterprises. This should not be a surprise to us: for assumptions (concealed or not) always have consequences.⁸

The structure of the paper will roughly be as follows: I begin by stating the MR contention for philosophy of mind and rehearse why MR poses a problem for type reduction. Philosophers have regarded MR as expressing either an empirical truth or some sort of conceptual necessity; I examine the cogency of these two claims in turn, concentrating on the claim of conceptual necessity. Along the way I also rehearse the intellectual history of MR and reductionist replies and note two varieties of MR: (1) MR across structure types (henceforth, type MR) and (2) MR within a token system (henceforth, token MR). The overriding goal of this paper is to dissect (1) the ontological commitments and (2) theories of causation of nonreductive physicalists and the type reductionists and to evaluate MR within different ontological frameworks. Disclaimers: In the past decade, a number of philosophers have argued that physicalism is not a well-formed thesis.⁹ I will assume for the sake of discussion that physicalism is a substantive thesis that is concerned with the laws and entities of an ideally completed physics. Nothing essential, however, rides on this assumption.¹⁰ I shall ignore worries about qualia, normativity and agency and will not discuss various alternative frameworks, such as substance dualism. My goal in this paper is to resolve the debate between current reductive and nonreductive physicalists, and so the argument can be read conditionally: *if physicalism is true, what is the most plausible view on the relationship of the mental to the physical?*

2. The Ontology of Multiple Realizability

2.1. MR, Intuitively

Multiple realizability in the philosophy of mind is the contention that a given mental kind (property, state, event)¹¹ can be realized by distinct physical kinds. Putnam was the first

⁷ This reminds me of Jerry Fodor's quip in his (1985, p. 76): "It rained for weeks and we were all *so* tired of ontology, but there didn't seem to be much else to do."

⁸ Compare John Heil's comments in his (2000): "I am convinced, however, that issues now at the forefront of the philosophy of mind are fundamentally metaphysical in character. (This is scarcely surprising. The philosophy of mind is, after all, a kind of *applied metaphysics* [emphasis not in original]. It is crazy to think that philosophers of mind can ignore or remain neutral on questions of ontology.)"

⁹ See Crane and Mellor (1990) for arguments to this conclusion.

¹⁰ I will also ignore empirical worries concerning the discontinuities between various varieties of MR. Polger (ms.) differentiates between MR in cases where the realizers are somewhat similar (say the realization of the visual system in higher primates) and where the realizers are radically distinct (say Martians and green cheese and silicon androids all realizing pain).

¹¹ [The ontology of these metaphysical entities is as yet an unsettled issue (Stewart, 1997).]

to publish this intuition. I once presented this thesis to my biologist girlfriend who promptly rejected it on the grounds that my intuitions were not well founded scientifically. She remained unconvinced after many minutes of hand waving on my part and many fantastic examples ranging from dogs to Martians to robosapiens. I will return to ask whether there is empirical evidence for MR but put this worry aside for the moment. A philosopher's (contentious) apology: philosophy proceeds with a radically different set of normative criteria from science. Intuitions, like MR, drive much philosophizing. What sustains them are sometimes bizarre philosophers' examples (intuition pumps).

Take pain for example (sorry!). It seems plausible that a whole range of animals, from octopuses through humans, are capable of feeling pain, i.e. they are reasonable candidates for pain-bearing types. But consider how vastly different the physiologies and anatomies of these organisms are. Pain, then, must be a multiply realizable kind. Now consider the genius neuroscientist Koch who invents the prosthetic neuron. To prove the efficacy of the Koch neuron, he asks his neurosurgeon colleague Crick to replace one neuron in his primary visual cortex (V1) each day with a Koch prosthetic neuron till his V1 consists entirely of Koch neurons. Koch's trusty graduate students and postdocs perform a daily battery of behavioral tests to see if all functions associated with V1 remain constant. The experiment proves to be a great success. So mental functions strongly correlated with V1 activation are multiply realizable. One can imagine Koch deciding to continue replacing neurons in other parts of his brain with Koch neurons because prosthetic neurons are less susceptible to atrophy due to advanced synthetic materials. He repeats this implantation process a couple times throughout his career, swapping the latest prosthetic neurons for the outdated models. These examples and many other twisted tales of multiply realized minds—the conscious silicon computer HAL, Martians, and other far-flung life forms from the furthest reaches of the universe—sustain the MR intuition.

It is now commonplace to assume that mental kinds are multiply realizable and that MR provided a decisive refutation of type physicalism. The type physicalist thesis is that mental states and processes are type identical to neural states and processes, i.e. there exists a necessary and sufficient physical condition, namely a yet-to-be-discovered neural kind, for the occurrence of every mental kind. On this picture, the relationship between mental and physical kinds is one-one. If MR is plausible, a given mental kind will have distinct physical realizers, i.e. one-many. Since any of the distinct physical realizers of a mental kind M are sufficient physical conditions for M 's occurrence, and none necessary, there are no necessary *and* sufficient physical conditions for the occurrence of mental kind M , contrary to the assertions of the type physicalist.¹²

2.2. MR, Empirically: Doubtful

There are standardly two ways of understanding the multiple realizability claim. Putnam saw MR as an empirical truth (one that falsified the empirical thesis of mind-brain identity), while later philosophers have regarded MR as expressing a conceptual

¹² See Fodor (1974).

necessity. I see myself as basically responding to the second construal of MR. It is seldom noted how difficult it is to establish MR as an empirical truth—the first construal—in the first place. There are two worries.¹³ One is the difficulty of individuating psychological kinds in empirical settings. Consider Ron Endicott's papers that have described how mental functions are sustained through neural plasticity, thereby suggesting the multiple realizability of mental states. But this can only be true on the grossest level of grandmother (folk psychological) individuation: “Look, he can still talk even though an iron stake went through his skull!”¹⁴ Furthermore, when we move to consider the difficulties of our projections of exactly the same mental states onto other primates and organisms, we see that the projections which sustain the MR intuition are “untestable” (though there must be some sort of continuity, for otherwise studies of macaque visual systems wouldn't help us understand ours at all). If indeed there is a strong continuity between neural and functional states across species with at least fairly sophisticated level of cognitive function—as evidenced by the reliance of neuroscientists on “psychological measures in mapping the brain and [their doing] so in a comparative fashion” successfully (Bechtel and Mundale 1997, 2000)—then it seems that some sort of identity thesis is far more likely, at least for organic life forms. Perhaps this result pressures us to return to the sort of species-specific-reduction reply that both Lewis and Kim suggest we give to the MR argument to irreducibility.¹⁵ But of course then one can complain that we are exacting an empirical criterion on MR, which is but a philosophical thesis. But remember, we are considering the first interpretation of MR, Putnam's interpretation, where MR is an empirical truth.

2.3. MR, Conceptually: Metaphysical Alternatives

2.3.1. MR under causal realism and the causal powers view of properties

Let us now turn to examine the interpretation of MR as conceptual necessity. If one adopts a sparse ontology—where the only properties admitted into the ontology are those that “cut nature at its joints”—and individuates properties on the basis of their causal powers,¹⁶ then MR amounts to no more than a second-order manner of picking out first-order entities. This is indeed the functionalist construal of mental properties.¹⁷ One must then ask what the ontological status of these second-order properties is.¹⁸ Let us now turn to examine the possible cases.

¹³ See Zangwill (1992) for further discussion.

¹⁴ This is, of course, a caricature—but one that contains a kernel of truth.

¹⁵ See pp. 233-236, Kim (1996), Kim's comments on the structure-restricted correlation thesis in his (1992), and Lewis (1969).

¹⁶ Shoemaker gives a robust account of this in his (1980), while Kim has often assumed a similar individuation criteria that he has christened “Alexander's dictum”: a property is real if and only if it has causal powers. Note that Kim's individuation criterion is weaker than Shoemaker's.

¹⁷ The relation between functionalism and multiple realizability is fuzzy. One can easily imagine examples of functionally defined kinds which are not multiply realizable. Consider a functional characterization of any elementary particle, in term of the causal-functional role in high energy physics interactions. Even though the elementary particle in question can be functionally defined, it is not multiply realizable.

¹⁸ Note that functional properties are second-order properties.

Epiphenomenalism. If second-order properties are epiphenomenal, as Block (1990, 1995) suggests, then mental properties contribute nothing *qua* causal and since properties are exhausted by their causal powers, there are no *effective* mental properties and the question of reducibility or irreducibility dissolves. All that exists are physical properties, since only these are causally potent. On this picture, mental predicates are linguistic locutions having positive epistemic status that play useful roles in the communication and cognition (folk psychology) of human cognizers, but no more.

Token identity. If second-order properties pick out psychological properties, and if they inherit the causal powers of their token physical realizers, then their causal powers are nothing over and above the causal powers of their token physical realizers. These second-order properties are then identical to the first-order realizers since we individuate properties based on their causal contribution. Furthermore, if the first-order physical realizers are distinct, as is standardly assumed, then they will be in each case distinct properties. There is a tension in holding the token identity of the mental property in question with each of its *distinct* physical realizers: imagine a mental property M , which is realized distinctly in human, crocodile and octopus neural circuits, N_h , N_c , and N_o respectively. If indeed the first-order physical realizers are distinct, it must be because each physical realizer contributes different causal powers, since we assume properties to be exhausted by their causal roles. So $N_h \neq N_c \neq N_o$. But $N_h = M$, $N_c = M$, and $N_o = M$, so by transitivity of “=”, $N_h = N_c = N_o$. To escape contradiction ($N_h \neq N_c \neq N_o$ and $N_h = N_c = N_o$), the only consistent way of interpreting the second-order mental property is as a predicate which does not figure in ontology, i.e. as an epistemic-linguistic locution that picks out first-order properties. But we can have as many of these linguistic entities as we want; jade is an excellent example.¹⁹

(Note the differences between Davidson-style and Kim-style events and difficulties with Davidsonian events. It is important to realize that one’s ontology of events can impact one’s understanding of token identity. In particular, differences between Davidson and Kim events are central to the issue of reducibility or irreducibility when one espouses a token identity relation. Kim understands events as property exemplifications at specific times, whilst Davidson understands events as concrete particulars that can fall under event-kinds. On Davidson’s account of events, token identity holds when an event falls under both a mental event-kind and a physical event-kind. On this account of token identity, the relationship between mental and physical properties is hopelessly obscure. Why so? Consider an object which has size and has shape. What is the relationship of size and shape? Is it one of identity? We know it is not. But in general, the relationship between two properties of an object may not be deduced simply from the fact that an object has both those properties. Analogously, even though a concrete event may fall under both mental and physical event-kinds, the relation of the mental and physical aspects of the event is unclear. What I think Davidson’s account amounts to is agnosticism with regard to the mind-body relation. Throughout this paper, I will be assuming Kim’s account of events. See pp. 58-62, Kim (1996) for a discussion of these issues. See also LePore and Loewer (1987).)

¹⁹ See Kim (1992), pp. 11-19.

Running the above line on the ontological status of second-order mental properties requires a story about properties of ordinary objects, for otherwise one would be vulnerable to a *reductio* claiming that the exact same worries about mental causation generalizes to anything which is not a fundamental physical property.²⁰ Here's a quick story: We are to understand the properties of complex physical states as being the composition of all²¹ their microphysical properties. In everyday experience, ordinary objects appear to have unique causal powers—e.g. a square peg with 1 inch sides cannot fit in a round hole with a 1 inch diameter—because the composition of microphysical properties results in apparently unique properties to human observers. However their properties *just are* the composition of the microphysical realizers (i.e. these are what Armstrong calls structural properties) and are nothing over and above the properties of their microphysical constituents. The apparent novelty of causal contribution is yet another epistemic artifact.

Nonstandard views of realization. We now turn to consider an alternative view of realization which purports to allow for MR and irreducibility. Sydney Shoemaker sets out his most recent view of realization in his (1999), wherein properties are individuated by their causal features.²² This is cashed out in terms of what he calls “forward-looking and backward-looking causal features” which allows for more fine-grained property individuation than the token identity view. The forward-looking causal features of a property *P* are the causal powers that are had by *P* (“... what contribution their instantiation can make to causing various effects”), while the backward-looking causal features are the set of causal features of states of affairs that might cause *P*'s instantiation. So on Shoemaker's picture, if property *P* realizes property *Q*, then the “forward-looking causal features of *Q* are a subset of the forward-looking causal features of *P*, and the backward-looking causal features of *P* are a subset of the backward-looking causal features of *Q*” (see diagrams 1 and 2). Notice that this allows that “it is sometimes the realized property, not its realizer, that causes a certain effect; this will be so when the causal features involved in the causal transaction belong to the subset which the realized property shares with all its realizers.”

So on Shoemaker's picture, a property *M* is multiply realizable and hence irreducible since *M*'s forward-looking and backward-looking are not identical with any of its distinct neural realizers, *N*₁, *N*₂, ..., *N*_n. Notice, however, that since properties are to be individuated on the basis of their forward-looking and backward-looking causal features, nothing prevents us from identifying property *M* in the diagrams below with causal features *P*₁, *P*₂, and *P*₃, which we identify as property *N*'. For example, we can construe property *N*₁ as *N*' plus *P*₄, and property *N*₂ as *N*' plus *P*₅. At this juncture, it seems that Shoemaker has two options: (1) dismiss *N*' as an empirically nonexistent property, or (2)

²⁰ See Kim (1997), Noordhof (1999), Kim (1999), Kim (1998), Block's (ms.), Loewer (2001), and Menzies (2001).

²¹ [Idea of “composition” needs to be stated more carefully. Ontological issues impinge: structural universals? tropes? or a linguistic view of compositionality (the whole is a function of its parts)?]

²² It is not immediately clear to me what causal features are. Are these causal powers? Are they finer grained? Which individuation criteria are we to contrast the causal features view against?

identify M with N' , and understand M as a second-order linguistic locution that picks out its distinct first-order realizers $\{N_1, N_2, \dots, N_n\}$. Because Shoemaker chooses to individuate properties on the basis of their causal features,²³ he may not reply to the reduction of M to $N' = \{P_1, P_2, P_3\}$ by claiming that the identification fails because N' does not exist empirically since N' just is the causal features P_1, P_2 , and P_3 .²⁴

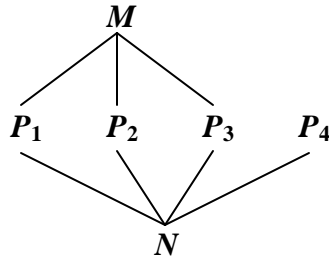


Diagram 1. A property N realizing M in Shoemaker's picture.

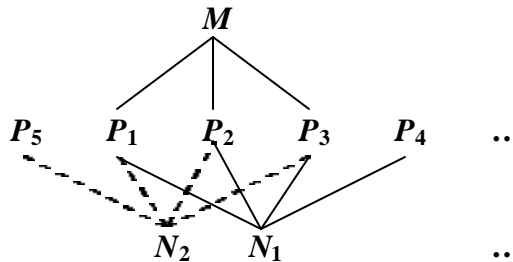


Diagram 2. Multiple realizability in Shoemaker's picture.

In all the cases above, the multiple realizability amounts to no more than a second-order linguistic locution that picks out or names the set of properties $\{P_1, P_2, \dots, P_n\}$.²⁵ This is certainly insufficient to argue for irreducibility. The question of reducibility does not even arise in these contexts, since the supposed mental properties fail to qualify for propertyhood. So the only way that there could be a relationship of irreducibility between mental properties and their physical realizers is if one conceals a property

²³ See Shoemaker 1980.

²⁴ I not here consider another view of realization, somewhat similar to Shoemaker's where instead of M 's causal features being a subset of M 's realizers', M 's realizers' causal features are a subset of M 's. Multiple realizability on this picture would look like this: assume M is realized by two distinct neural properties N_1 and N_2 . $M = \{P_1, P_2, P_3, P_4, P_5, P_6\}$, $N_1 = \{P_1, P_2, P_3, P_4\}$, and $N_2 = \{P_1, P_2, P_3, P_5\}$. Now, M is not reducible to either of its realizers, but someone who espouses such a view would have to explain how P_6 emerges mysteriously. My treatment of Shoemaker's picture in this section is heavily indebted to Heil's treatment in his (1999).

²⁵ In his (1999), Kim arrives at a similar conclusion. He suggests that we "give up [the higher-order mental property] E as a genuine property, only recognizing the expression " E " or the concept of E . As it turns out, many different properties are picked out by E , depending on the circumstances One could argue that by forming "second-order" functional expressions by existentially quantifying over "first-order" properties, we cannot be generating new properties, only new ways of indifferently picking out first-order properties in terms of certain causal specifications that are of interest to us (p. 17)." See also Heil (1999) and Antony (1999), pp. 1-9.

dualism to begin with; the multiple realizability argument to irreducibility is thus circular. Functionalism (the NRP version) then fails to be established. Functional kinds can however be identified with their physical realizers as in Lewis-Armstrong causal functionalism.²⁶ This is just as in the carburetor case. It is not the property of being a carburetor that does the causal work. It is the internal workings of the X, Y and Z parts of the carburetor and their interactions that allow it to play the causal role that is required. One can call those things “carburetors” if one likes.

2.4. MR under a deflationary account of causation

Turn now to consider multiple realizability and nonreductive physicalism under a broadly Humean, deflationary perspective toward causation. Hume’s view is that there is nothing intrinsically intelligible about causality, and that there are no necessary connections between a cause and its effect. “The contrary of every matter of fact is still possible; because it can never imply a contradiction” (Hume 1748, p. 25). Our innate intuitions concerning intrinsic necessary connections existing between causes and effects are due to our recognizing regularities in our experience where the members of a class of events and objects are constantly conjoined with members of another class of events and objects and our projecting an *intrinsicness* into the causal relation. Hume further suggested that the truthmakers of purported causal relations are the associated counterfactual conditional statements. He wrote: “We may define a cause to be *an object, followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*. Or, in other words, *if the first object had not been, the second never had existed*” (Hume 1748, P. 76). Hume did not go on to develop a counterfactual analysis of causation, and it was not until the 1970s that counterfactual treatments—largely owing to the seminal work of David Lewis²⁷—became immensely popular.²⁸

It turns out that under a Lewisian counterfactual analysis of causation, the causal efficacy of higher-level properties is not an issue at all (and we should expect overdetermination). Proponents of nonreductive physicalism standardly appeal to counterfactual accounts of causation where every cause/effect pair instantiates a law. NRP defenders typically also espouse a rich (almost profligate) ontology, where any predicate that is subsumed under a law (regardless of whether the law is a strict law or a *ceteris paribus* one) is a causally potent property.²⁹ Fodor’s line of thought on the predicate property distinction is that any

²⁶ I am not claiming here that Lewis or Armstrong’s metaphysical views and their philosophies of mind are indeed consistent. This is not true, especially in the case of Lewis. Lewis is a Humean about causation, and as we shall soon see, there is a great tension between his causal functionalism and his Humean counterfactual treatment of causation. Terry Horgan recently noted this point in his forthcoming essay “Multiple Reference, Multiple Realization, and the Reduction of Mind”.

²⁷ See Lewis (1973).

²⁸ Hume actually was confusing two disparate deflationary accounts of causation, the regularity theory and the counterfactual analysis.

²⁹ I.e. every predicate that figures in a covering law explanation corresponds to a property. There is a further implicit assumption that any explanatorily relevant property is also causally relevant. Epiphenomenal properties are explanatorily irrelevant and hence do not figure in the nonreductive physicalist’s account.

predicate that is subsumable under law exists and is a causally potent property. He charges Kim with making an unnecessary constraint on properties: that they must be subsumable under physical law, which is why higher-order properties on Kim's count turn out to be causally impotent. Fodor himself is agnostic on metaphysical issues³⁰ but has acknowledged that if one adopts a Humean attitude toward causation, many of the worries evaporate.³¹ The counterfactual defense of NRP has been undertaken by numerous philosophers who espouse some sort of Humean supervenience of laws on the set of local, intrinsic facts, including Block (1990), Horgan (1989, 1993, and forthcoming), LePore and Loewer (1987, 1989), and Loewer (forthcoming). Among these, Horgan's account is the most carefully worked out one, and I will sketch his defense against arguments for the causal exclusion of higher-order mental properties below. Horgan's account, like all other defenses of NRP, relies heavily on the notion of explanatory relevance. Horgan begins in his (1993) by giving a careful account of causal explanation. Consider the case where we want to give an explanation of the causal transaction between a phenomenon c and a phenomenon e , where c is instantiating a property of type C and e is instantiating a property of type E . For properties C and E to be genuinely explanatorily relevant to the causal transaction between phenomena c and e , Horgan, following Woodward (1979), requires not only that c caused e and that c and e are subsumable under a counterfactual, law-like generalization (this satisfies the Fodor propertyhood requirement), but importantly also that C and E must fit into a suitably rich pattern of counterfactual relations among properties. Horgan calls this the counterfactual pattern conception (CP) to explanatory relevance.

Under the CP conception, a single phenomenon, e.g. that c caused e , can be subject to a variety of equally valid explanations involving different theoretical levels of counterfactual, law-like generalizations. Typically, the level of theoretical explanation is discourse sensitive and determined by the relevant level of detail. Ignoring worries about the *ceteris paribus* nature of psychological laws,³² each level of counterfactual generalizations is equally explanatorily valid for the context of discourse in question. Consider a certain counterfactual generalization about neurophysiological states that relates a state N_1 to a state N_2 . Even if multiple realizability doesn't hold, explanation in terms of the conjunction of huge numbers of quantum mechanical states that make up complex molecular structures that in turn make up the neurophysiological states seems infeasible. Certainly, just churning the numbers out (most likely with the help of significant computational aids) would suffice for an instrumentalist (or behaviorist)

³⁰ Fodor is apathetic about physicalism almost to the point of agnosticism. He writes in his "Making Mind Matter More": "I'm not really convinced that it matters very much whether the mental is physical; still less that it matters very much whether we can prove that it is. Whereas, if it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my saying ... if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world." (Fodor, 1989, p. 77)

³¹ Fodor, in conversation.

³² The claim that the lower-level covering laws of physics are "strict" and "exceptionless" is quite preposterous. Quite the contrary, physical laws, like Newton's laws and Hooke's law hold only *ceteris paribus* as well. Newton's laws for example only hold at velocities radically smaller than the speed of light. Having said that, we expect the fundamental field equations (or string theoretic equations for that matter) of completed physics to be exceptionless.

analysis, but substituting the vast quantum-state conjunctions for a neurophysiological predicate in a covering law would scarcely be “explanatorily adequate”. The law wouldn’t make any sense. So at each grain of analysis, there is an explanatorily appropriate level of counterfactual, law-like generalizations (which, as you might remember, are embedded in some rich network of counterfactual dependencies). Remember also that any predicate that figures in a counterfactual, law-like generalization names a causally efficacious property, so Horgan’s Humean account gives us a rich, layered cake of properties.

3. Humean Worries: Irrealism, Inter-Level Constraints and Ontic Primacy

An immediate objection against Horgan’s account is that explanation is an anthropocentric epistemological notion, and fails to speak to the ontological worries at hand. Horgan is aware of this objection,³³ but fails to grasp the full extent of the ontological problem. In fact, he makes a fatal error: he implicitly conflates explanatory relevance with causal relevance. After his explication of the CP conception, Horgan rushes to defend CP against what Jaegwon Kim calls *explanatory irrealism*, which Kim describes as “the view that the relation of being an explanans for, as it relates *C* and *E* within our epistemic corpus, is not, and need not be, ‘grounded’ in any objective relation between events *c* and *e*”.³⁴ Defending the Humean account against Kim’s explanatory irrealism charge is easy, as the irrealism charge is far too strong. Horgan’s response to Kim’s charge is appropriate: indeed an advocate of the CP conception can point to objective grounding for the explanans/explanandum relation, for on the NRP count, causal closure of physics implies that for every event there is a complete causal explanation in terms of physical properties. The nonreductive physicalist then tells a story where the complete physical causal story doesn’t trump counterfactual, law-like causal generalizations at other levels of explanation, but instead provides the ontological grounding for all (remember that the nonreductive physicalist espouses a position of token identity with respect to the mental-physical relation^{35, 36}) other levels of causal

³³ Horgan writes: “This kind of context/purpose relativity is entirely compatible, as far as I can see, with the contention that facts about explanatory relations are objective and mind-independent. Hence the CP conception does not embody any commitment to what Jaegwon Kim calls *explanatory irrealism*.” (Horgan 1993, p. 300)

³⁴ Kim (1988), pp. 226-227.

³⁵ In general, the nonreductive physicalists espouse token identity as the relation between physical properties and any other properties from higher levels in the layered NRP picture of the world. There are also other varieties of nonreductive physicalism where the relationship between higher-order properties and physical properties is not one of token identity, but global supervenience. Such accounts are often motivated by concerns regarding wide content or externalism. Note also that the nonreductive physicalist espousing such a global supervenience account needs to give an account of what it means for a set of properties to globally supervene on another set of properties beyond just stating definitions characterizing some sort of property-dependent modal covariation. See McLaughlin (1995) for an exhaustive treatment of supervenience. See Horgan (1993) and Kim (1998) for arguments to the conclusion that supervenience physicalism fails to characterize a distinctive theory of mind, and that consumers of supervenience need to further explain the property-dependent modal covariation.

³⁶ At the moment I simply want to note that there is a tension in the token identity view. Giving physical properties the pride of place in our ontology, whilst maintaining that properties at other levels in the layered ontology have separate ontological status, leads to numerous problems. These I will discuss soon enough.

explanation. However, Horgan still needs to respond to the charges that higher level explanatory relations are neither objective nor mind-independent.³⁷ Remember that the notion of explanation is “making sense of”—it is a cognitive notion and hence is relative to a class of cognizers—and that the relevance and efficacy of explanation is discourse and grain dependent. He has given us no reason to think the explanatory relevance implies causal relevance. His assertion that predicates figuring in explanations automatically acquire propertyhood is suspect; the tail is wagging the dog. What the arguments of Horgan (and Putnam³⁸ and Fodor³⁹ before him) effectively establish, is only that certain ways of characterizing reality are essential in helping human beings to make sense of the world, and these high-level vocabularies are, for all practical purposes, ineliminable.

There are further worries about the NRP metaphysical framework. On the CP conception, a single phenomenon, e.g. that *c* caused *e*, can be subject to a variety of equally valid explanations involving different theoretical levels of counterfactual, law-like generalizations. Furthermore, these nomic counterfactual generalizations are embedded in a larger complex network of nomic counterfactual generalizations that is the “web of science”, to use a Quinean term. We might ask what the interlevel constraints on this network of counterfactual levels are. We might think that for nonreductive physicalists the levels of nomic counterfactual generalizations are equally valid explanatorily, and hence properties at different levels of the layered ontology that NRP espouses have equal status. There is no question of ontic primacy of the physical; physical and mental properties are ontologically equally entities. To remedy this situation, NRP adds the assumption of causal closure. We might also ask what lines up the counterfactuals at different levels of explanation perfectly, such that physical, chemical, biological and mental explanations are equally valid and compatible. The story that Horgan provides us with is that everything else synchronically supervenes on the physical. But notice that unless NRP is espousing some sort of Leibnizian pre-established harmony (which nonreductive physicalists certainly don’t want) and unless Horgan is conflating explanatory relevance with causal relevance (or epistemological relevance with ontological relevance), the process of gathering nomic counterfactual generalizations (also known as “doing science”) is a messy diachronic process and it is by no means clear that counterfactual generalizations across levels will *always* be mutually consistent. There are two complaints that I have: (1) NRP seems to carve up scientific levels of explanation too neatly and discretely, but in reality the levels of explanation are continuous, e.g. chemical physics and physical chemistry, or biochemistry and molecular biology. (2) In fact, there is also good reason to doubt the autonomy of the upper ontological strata on the NRP picture because there are interlevel constraints that involve the day-to-day process of doing science. Ask any good scientist and he or she will tell you that each level of scientific explanation must be consistent with its adjacent levels. Thus, there must be some interlevel traffic, which is what allows knowledge acquired in

³⁷ He needs to confer propertyhood to properties at higher levels in the layered ontology, while at the same time grounding them in the physical realm.

³⁸ See Putnam’s “Philosophy and our Mental Life” (1973), in *Mind, Language and Reality* (1975), pp. 291-303.

³⁹ See Fodor (1974).

physics to filter up to molecular biology. A good example of this is how advancements in crystallography allowed for the elucidation of DNA structure. This filtering of knowledge is not transitive *qua* cognizers (an evolutionary biologist will *not* care about advancements in atomic theory), but since each level *must* be consistent with its adjacent levels (local consistency), knowledge acquired filters up the levels of explanation (there is a pressure toward global consistency).⁴⁰

As I noted earlier, under the CP conception, each level of explanation is *equally valid*, hence predicates that figure at different levels of explanation are *ontologically equal*. The ontological primacy of the physical that is so much a part of the physicalist picture does not figure in the CP conception as such and needs to be further stipulated; I find this somewhat unnatural. I would imagine that for physicalists (the “P” of the NRP) physical properties should have pride of place in their ontology and this ontological primacy should be a natural part of their metaphysical framework. Typically, the ontic primacy of the physical is hidden either in the realization relation, or some further stipulation like Lewis’ notion of naturalness. For example, Loewer and LePore in their (1987) require that an event *e*’s being *X* must explain *e*’s being *Y* if *e*’s being *X* is to realize *e*’s being *Y*. This explanatory requirement is meant to be stronger than plain physical necessity. [I find the realization relation somewhat suspect. Realization talk in the philosophy of mind was established in the 1960s by a series of early Putnam papers on the strength of computational analogies (abstract mathematical machines are realized by concrete physical devices). It has been noted more than once that the notion of realization has largely been taken for granted, and little or no work has been done to explicate the notion with regard to the more traditional options on the mind-body problem.⁴¹ Furthermore, it is not even clear that we understand the realization relation in the case of computers.⁴²]

4. Type and Token MR

In his forthcoming paper on “Multiple Realizability, Multiple Reference and the Reduction of Mind”, Horgan charges that David Lewis’ reductive causal functionalist account—which my account has broad sympathy with—fails to accommodate what Horgan terms strong multiple realizability. Horgan’s objection rests on a technicality based on Lewis’ use of the notion of a creature-kind in his explication of mental concepts as nonrigid designators. It is standard in the philosophy of mind literature to differentiate between two varieties of MR: type and token (my terms). Type MR is the idea that a mental property is realized by distinct structure types; token MR, which has been called radical MR (Polger) or strong MR (Horgan) in the literature, is the idea that a mental property might be realized by distinct physical tokens in a token system across times. As

⁴⁰ Consider the possibility of interlevel counterfactual generalizations.

⁴¹ See Kim (1998) pp. 7-9, Shoemaker (1999), and Block’s discussion of the realization relation in his (1995).

⁴² [Brian Cantwell Smith has repeatedly emphasized this point to me in conversation. According to Smith, the mind-body problem for computers is not trivial, and neither is the concrete/abstract distinction; one does not get notions like realization and the concrete/abstract distinction for free from the realm of computing. Philosophical analysis has yet to be performed on such basic notions in computing such as implementation and realization. See Sloman (1998) and Smith (1996).]

Horgan points out, Lewis' treatment of mental concepts as nonrigid designators having multiple referents flounders because he implicitly restricts the reference of mental terms to a creature-kind (remember the Lewis-Kim species-specific reduction reply). This is not a problem for Lewis' ontology, for Lewis can drop the restriction of reference being restricted to creature-kinds and further narrow the reference to token systems at specific times. One might complain that in making this move psychological predicates lose their generality and this poses problems for concocting psychological theories. But remember that we are concerned with the ontological status of mental predicates, not their epistemological status or utility in human discourse.

5. What Remains

The metaphysical groundwork and argumentative details in this paper are woefully incomplete at best, but I hope I have shown the importance of ontological candor in theorizing about the mind and provided the skeleton of arguments that should raise grave doubts about the multiple realizability argument to irreducibility qua ontology. I have no quarrel with the need for higher level vocabularies (with positive epistemic status) that pick out lower level realizers in our ongoing pursuit for knowledge about the world. In fact, if Putnam and Fodor's antireductionist arguments have shown anything, they have shown that such higher level vocabularies, such as mental discourse, are practically ineliminable. But if my arguments are sound, then these predicates fail to pick out fundamental properties. Perhaps this indicates that the only way to be a mental realist is to renounce our physicalist commitments and embrace some sort of robust dualism, but that remains to be seen.⁴³

⁴³ I am indebted to Jerry Fodor, Barry Loewer, and especially Timothy O'Connor and Brian Cantwell Smith.

References

- Louise Antony (1999), "Multiple Realizability, Projectibility, and the Reality of Mental Properties", *Philosophical Topics*, 26, pp. 1-24.
- William Bechtel and Jennifer Mundale (1997), "Multiple Realizability Revisited", *Proceedings of the Australian Cognitive Science Society*, URL = <http://artsci.wustl.edu/~bill/multiple.htm>
- (1999), "Multiple realizability revisited: Linking cognitive and neural states", *Philosophy of Science*, 66, pp. 175-207.
- John Bickle (1998), "Multiple Realizability", *The Stanford Encyclopedia of Philosophy* (Spring 2001 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/entries/multiple-realizability/>
- Ned Block (1990), "Can the Mind Change the World?", in Boolos (ed.), *Mind and Method*, Cambridge: Cambridge University Press, pp. zz-zz.
- (1995), "What is Functionalism?", in Guttenplan (ed.), *A Companion to the Philosophy of Mind*, pp. xxx-xxx.
- (ms.), "Do Causal Powers Drain Away?", manuscript.
- Tim Crane and D. H. Mellor (1990), "There is No Question of Physicalism", *Mind*, 99, pp. 185-206.
- Jerry Fodor (1974), "Special Sciences, or The Disunity of Science as a Working Hypothesis", *Synthese*, 28, pp. 97-115.
- (1985), "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum", *Mind*, 94:373, pp. 76-100.
- (1989), "Making Mind Matter More", *Philosophical Topics*, 17, pp. 59-79..
- John Heil (1999), "Multiple Realizability", *American Philosophical Quarterly*, 36:3, pp. 189-208.
- (2000), "Metaphysics of Mind", *A Field Guide to the Philosophy of Mind*, Nani and Marraffa (eds.), URL = <http://www.uniroma3.it/kant/field/mm.htm>
- John Heil and C. B. Martin (1999), "The Ontological Turn", *Midwest Studies in Philosophy*, 23, pp. 34-60.
- Terry Horgan (1993), "Nonreductive Materialism and the Explanatory Autonomy of Psychology", in Wagner and Warner (eds.), *Naturalism: A Critical Appraisal*.
- (1996), "Kim on the Mind-Body Problem", *British Journal for the Philosophy of Science*, 47, pp. 579-607.
- (2001), "Multiple Reference, Multiple Realization, and the Reduction of Mind", forthcoming in Sibel and Preyer (eds.), *Reality and Humean Supervenience: Essays on the Philosophy of David Lewis*.
- David Hume (1748), *Enquiry Concerning Human Understanding*, Selby-Bigge (ed.), Nidditch (rev.), 3rd edn, Oxford: Oxford University Press, 1986.
- Jaegwon Kim (1988), "Explanatory realism, Causal Realism, and Explanatory Exclusion", *Midwest Studies in Philosophy*, 12, pp. 225-240.
- (1992), "Multiple Realization and the Metaphysics of Reduction", *Philosophy and Phenomenological Research*, 52, pp. 1-26.
- (1993), *Supervenience and Mind*, Cambridge: Cambridge University Press.
- (1996), *Philosophy of Mind*, Boulder: Westview Press.

- (1997), “Does the Problem of Mental Causation Generalize?”, *Proceedings of the Aristotelian Society*, 97, pp. 281-297.
- (1998), *Mind in a Physical World*, Cambridge: MIT Press.
- (1999), “Making Sense of Emergence”, *Philosophical Studies*, 95, pp. 3-36.
- David Lewis (1969), “Review of Putnam”, reprinted in Block (ed.), *Readings in the Philosophy of Psychology*, vol 1, pp. 232-233, 1980.
- (1973), “Causation”, *Journal of Philosophy*, 70, pp. 556-567.
- (1995), “Lewis, David: Reduction of Mind”, in Guttenplan (ed.), *A Companion to the Philosophy of Mind*, pp. 412-431.
- Ernst LePore and Barry Loewer (1987), “Mind Matters”, *Journal of Philosophy*, 84, pp. 630-642.
- (1989), “More on Making Mind Matter”, *Philosophical Topics*, 17, pp.175-191.
- Barry Loewer (2001), “Review of Kim’s *Mind in a Physical World*”, manuscript, forthcoming in *Philosophy and Phenomenological Research*.
- Brian P. McLaughlin (1995), “Varieties of Supervenience”, in *Supervenience: New Essays*, Savellos and Yalçin (eds.), pp. 16-59, Cambridge: Cambridge University Press.
- Peter Menzies (2001), “The Causal Efficacy of Mental States”, in Monnoyer (ed.), *The Structure of the World: the Renewal of Metaphysics in the Australian School*, Vrin Publishers.
- Thomas Polger (2000), “Putnam’s Intuition and Multiple Realizability”, manuscript.
- Hilary Putnam (1973), “Philosophy and our Mental Life”, in *Mind, Language and Reality* (1975), pp. 291-303, Cambridge: Cambridge University Press.
- David Robb and John Heil (1998), “Ontology and Mental Properties”, manuscript.
- Sydney Shoemaker (1980), “Causality and Properties”, in van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor*, pp.109-136, Reidel: Dordrecht.
- (1999), “Realization and Mental Causation” in Elevitch (ed.), *Proceedings of the 20th World Congress of Philosophy*, Vol. IX: Philosophy of Mind, Bowling Green: Philosophy Documentation Center, Bowling Green State University.
- Aaron Sloman (1998), “Supervenience and Implementation: Virtual and Physical Machines”, Virtual and Physical Machines Technical Report, School of Computer Science, University of Birmingham.
- Brian Cantwell Smith (1996), *On the Origin of Objects*, Cambridge: MIT Press.
- Helen Stewart (1997), *The Ontology Of Mind: Events, Processes, and States*, Clarendon Oxford Press: New York.
- James Tomberlin (ed.) (1997), *Philosophical Perspectives*, Vol. 11: Mind, Causation, and World.
- Peter van Inwagen (ed.) (1980), *Time and Cause: Essays Presented to Richard Taylor*, Reidel: Dordrecht.
- James Woodward, (1979), “Scientific Explanation”, *British Journal for the Philosophy of Science*, 30:41-67, pp.54-55.
- Nick Zangwill (1992), “Variable Realization: Not Proven”, *The Philosophical Quarterly*, 42:167, pp. 214-219.