

Multiple Realizability, Qualia and Natural Kinds

Andrew Bailey

Department of Philosophy

The University of Guelph

Guelph, Ontario N1G 2W1

abailey@uoguelph.ca

(519) 824-4120 x53227

Multiple Realizability, Qualia and Natural Kinds

Are qualia natural kinds? In order to give this question slightly more focus, and to show why it might be an interesting question, let me begin by saying a little about what I take qualia to be, and what natural kinds. For the purposes of this paper, I shall be assuming a fairly full-blooded kind of phenomenal realism about qualia: qualia, thus, include the qualitative painfulness of pain (rather than merely the functional specification of pain states), the qualitative redness in the visual field that typically accompanies red discriminations, the taste of lemon (independently of the fact that such states are normally caused by lemons and give rise to puckering of the lips, etc.), and so on. In other words, I am assuming the falsity of functionalism with respect to qualia, though I am not for a moment assuming dualism.

Natural kinds are standardly thought of as groups of objects that have some theoretically significant property, or properties, in common. These properties—sometimes thought of as essential properties—are theoretically significant because an object's membership of a given natural kind determines how that object will behave, in what set of interactions it can figure, and what further properties it can or cannot acquire. Thus, natural kinds can form a system that support the explanation and prediction of the behaviour of those objects, in accordance with lawlike regularities: in other words, natural kinds can underpin—provide a basis for—inductive reasoning. Common examples of natural kinds include biological species such as rabbits, elms and whales; chemical elements and compounds such as oxygen, carbon and H_2O ; and stuffs such as salt, wool and heat in a gas.

Something is a natural kind, then, roughly if it is a *scientifically respectable* kind: if it is a kind whose members all have some kind of essential similarity, and form part of a natural scientific domain governed by laws and regularities that support, for example, inference to counterfactual cases, and so on. It seems, then, that the project of showing that qualia are natural kinds is also the project of showing that qualia are scientifically respectable and causally significant.

It is also worth mentioning at this point that, though this is not something I shall stress much in this essay, it is part of the modern view of natural kinds that their particular nature is not discoverable *a priori*: for example, what makes water water—its chemical composition—cannot be known merely from consideration, or stipulation, of the concept *water*, but must be discovered through empirical investigation. Hence, if qualia *are* natural kinds, it is likely to follow that their nature is discoverable only *a posteriori* (rather than, say, through analysis of the meanings of certain

psychological terms).

So, now, are qualia natural kinds? What considerations might count for or against the suggestion that they are? There are, of course, various issues here—starting, perhaps, with brute intuitions about the scientific respectability or otherwise of phenomenal ‘raw feels’—but one of the main ones, it seems to me, has to do with the phenomenon of multiple realizability. Recall the following canonical piece of reasoning from Hilary Putnam’s seminal article:

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 must *not* be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extra-terrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may *be* pain. (Putnam 1967/1975, 436)

As Stephen Yablo succinctly puts it, “[p]roperties are identical only if each necessitates the other; but any physical property specific enough to necessitate a mental property is inevitably *so* specific that the converse necessitation fails” (Yablo 1992, 250).

Why is multiple realizability a problem for natural kindhood? It is a problem, on the face of it, because it shows that the similarities among certain ‘manifest’ kinds—such as pain, desiring *p*, being a flying thing, or being a spotted thing—are not supported by underlying, scientifically interesting commonalities. Chemical compounds and biological species are natural kinds *in virtue of* their underlying, empirically discovered natures (their chemical composition or, perhaps, their shared evolutionary history). It is these natures that underpin inductive generalizations over the members of those kinds. If water had turned out to be one chemical compound in the Pacific, another in the Atlantic, another in the Great Lakes, and so on, then it would not have been a natural kind. *Being pale blue and smaller than a breadbox* is not a natural kind just because—apart from these two properties—there is no underlying, scientifically significant set of properties that all the objects that satisfy that description have in common. In exactly the same way, as Kim puts it, “...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 , ... such that N_h realizes pain in humans, N_r realizes

pain in reptiles, N_m realizes pain in Martians, etc.” (Kim 1992, 5).

I am going to take it, then, that multiple realizability is inconsistent with natural kindhood. In fact, the situation may be rather more complicated than this. For example, it might be said that there are other varieties of natural kinds than *physical* kinds—in particular, there might possibly be *functional* natural kinds, and if so the members of such kinds would not be expected to have similar physical compositions but instead, perhaps, relevantly similar detailed functional roles. Similarly, the members of these ‘higher-level’ natural kinds might be said to form domains that are subject to ‘higher-level’ natural laws, such as perhaps the laws of psychology. On this way of thinking, multiple realizability does not call into question the natural kindhood of these kinds—and so, does not in this way call into question their scientific respectability—but it does introduce the familiar, vexed question of the relationship between the ‘higher-level’ laws thus introduced and those of physics, and hence the equally familiar and equally vexed problem of mental causation (and so by this indirect route still throws doubt on the scientific status of the higher-level kinds).

I, however, announced at the outset that I am treating qualia, not as functional kinds, but as—as it were—phenomenal kinds, realized by the brain. On this view pain, for example, is perhaps best thought of—albeit loosely—as a kind of ‘stuff,’ like heat or salt, whose essential micro-properties explain both its macro-properties and its lawlike relations to other objects. Thus, although on the one hand multiple realizability is, at least on the face of it, a more pronounced phenomenon for functional properties than for phenomenal ones, on the other the functionalist response to the multiple realizability threat to natural kindhood is not available to us. (Some of what I shall have to say about multiple realizability in a moment, however, will call into question the whole strategy of appealing to a distinction between ‘functional’ and ‘physical’ levels of properties and laws, so perhaps this is not quite the handicap it seems.)

So, let us take it, if qualia as I characterize them are multiply realizable then they are not natural kinds.¹ Recently, there have been a number of articles—by John Heil, Stephen Yablo and David Robb—exploring the prospect of understanding multiple realizability as a version of the *determinable/determinate* relation. That is, in bare summary, instead of thinking of multiple realizability as a relation between higher-level and lower-level kinds or properties, we are to think of it as analogous to the relation between, for example, colours and their particular shades—say, redness

¹ Structure-restricted qualia—such as human pain or octopod pain—might perhaps still be natural kinds (see Richardson 1979, Kim 1992), but qualia *sans phrase* could not be and it is the latter question that we are pursuing here. As Ned Block puts it, species-restricted reductionism “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’” (1980, ?).

and crimson, scarlet, magenta or rose—or between dancing and engaging in particular kinds of dance, such as the waltz or the tarantella. I shall explain this model of multiple realizability, and then consider its implications for the status of qualia as natural kinds. I will focus on the presentation of the model in John Heil’s 1999 article “Multiple Realizability.”

It is standard to take multiple realizability to be a relation among properties at different levels, whereby a higher-level property is multiply realizable if it can be realized by more than one lower-level property. As John Heil points out this has as a consequence (or a presupposition) the ontological position that “the world is ‘layered.’ Reality comprises ‘levels’ of objects, properties, and laws” (Heil 1999, 189). To be multiply realizable is, of course, to be realizable. If some higher-level property—pain, say—is realized in a human being by the firing of those perennial c-fibres, then it seems that *two* properties are being tokened at the same time: there is the property of being a c-fibre firing, to be sure, but there is *also* the *distinct* property of being a pain. This higher-level property of being a pain figures in its own right in higher-level laws (in this case, psychological laws), and features in higher-level explanations.

Heil argues that this multi-level account of multiple realizability is misconceived. It invites the following problem: why should we suppose that the higher-level property, in each of its tokenings, is *distinct* from the lower-level property that realizes it on that occasion? Suppose, for simplicity, that pain is realizable by just three lower-level properties: call them mammal-neural, mollusc-neural, and martian-hydraulic. Why should we suppose that the world contains instances of those three properties *plus* instances of a fourth property, pain?

The standard sort of reason for this insistence, Heil plausibly suggests, is the sentiment that the special sciences are *autonomous*: they capture laws and generalizations that are not replaceable by those of lower-level sciences (such as physics), and hence the entities that figure in those laws and generalizations are similarly autonomous. Since the entity in question is a property, this in turn, suggests the following conception of properties: “[a] property contributes uniquely to the ‘causal powers’ of objects possessing it” (Heil 1999, 192). And it follows from *this* that “[i]f property *A* and property *B* affect, or would affect, the causal powers of objects in precisely the same way, then *A* and *B* are the selfsame property” (Heil 1999, 193).

But now, if some mental property *M* is to be a distinct property from its various neural realizations N_1 , N_2 , and N_3 , it must have different causal powers than those neural properties. But how is *this* to be made sense of?

What is hard to see ... is how *M* could bestow causal powers on *a* other than those bestowed

by N_1 . If we take N_1 to be doing M 's work, then in what sense could *both* M and N_1 be present in a simultaneously? (Heil 1999, 193)

Heil sketches two possibilities. In one, the causal powers of M make up a *subset* of those of each of its realizers; in this case N_1 will have all the causal powers of M plus others as well. There is much that is attractive about this picture, but the central problem still remains: in what sense does an agent a , who possesses N_1 , *also* possess M ? What *more* is there to a 's possessing M than its possessing N_1 ? There is nothing more, unless we engage in a kind of 'double-counting' of causal powers, counting them once when they are bestowed by N_1 , and again as being bestowed by M .²

Perhaps, though, we could say that N_1 , N_2 , and N_3 are complex properties, and that—though they differ in other ways—they each have the property M as a constituent? “But then it is hard to see why mental properties are not straightforwardly identifiable with material properties, those properties, namely, that realize their distinctive causal powers” (Heil 1999, 195). Furthermore, this prospect is precisely as implausible as the original intuitions behind multiple realizability are potent: “how could M , with its distinctive intrinsic nature, be a constituent of N_1 , a constituent of N_2 , and a constituent of N_3 , when each of these— N_1 , N_2 , and N_3 —differs dramatically from the others?” (Heil 1999, 198).

The second prospect for multiple realization as a relation between different properties that Heil canvasses is the proposal that the causal powers of M might be a *superset* of the causal powers of each of its individual realizers. One obvious disadvantage of this tack is that it makes the realization relation completely mysterious: where do these extra causal powers of M come from, if not from M 's physical realization? Furthermore, we encounter a sort of inverse of the problem we saw in the case where M 's causal powers are a subset of N_1 . In that case, it seemed that the property M was subsumed by the property N_1 . Conversely, if the causal powers of each of N_1 , N_2 , and N_3 are subsets of the causal powers of M , it is not clear why a —in virtue of possessing M —does not also possess each of its supposedly disjunctive physical realizations.

What if M 's causal powers, instead of being a superset of the causal powers of its realizers, instead *overlapped* with those causal powers, such that M included some but not all of the causal powers of its realizers as well as additional causal powers that each of those realizers lack. This would solve the problem, but would still leave us with the mystery of how N_1 could realize M while lacking some of M 's causal powers.

² It will not do, Heil reminds us, to think of M as a sort of 'subset' of N_1 : property instances are concrete features of the world, not abstract relations like set membership.

Perhaps, then, what has gone wrong was our original supposition that for two properties to differ they must differ in their causal powers. Perhaps instead, we should think of mental properties as being *second-order* properties—pain, for example, might be the second-order property, M , of possessing a first-order property, N_1 , that occupies an appropriate functional role. Does this solve our problem? It seems not: the puzzle remains, what exactly would it be for a to possess M beyond a 's possessing N_1 . Talk of 'second-order' properties introduces nothing new into the mix.

Heil therefore introduces a quite different way of thinking of multiple realizability, one which does not suffer from the drawbacks of those just mooted. This approach is modelled, not on relations between properties at two different levels, but on the distinction between determinables and determinates: a mental property is to its physical realizations as a determinable, like redness, is to its determinates, such as scarlet or crimson. How does this solve the mystery of the apparent doubling-up of properties? After all, it is just as odd, and odd in the same ways, to say that a rubber ball possesses both the property of redness *and* the property of being scarlet (in the same way as it could be both white and spherical). Heil's response is, essentially, *to deny that determinables are properties*. As he puts it:

Many predicates apply to objects in virtue of properties possessed by those objects. Of these predicates, some designate properties shared by objects to which they apply [non-determinables, such as sphericity]. Others do not. Realism about a given predicate, ' Φ ,' requires only that ' Φ ' applies truly to objects in virtue of properties actually possessed by those objects. Realism does not require that ' Φ ' designate a property shared by every object to which it truly applies. If ' Φ ' does designate a property, then objects satisfying ' Φ ' must be identical (or exactly similar) in some respect, a respect in virtue of which ' Φ ' holds true of them. The bulk of our predicates, however, are satisfied by ranges of similar, but not exactly similar properties. (Heil 1999, 200–201)

On this model, we can say that the predicate *is in pain*, though it holds of particular objects at particular times, and does so in virtue of their properties, nevertheless does not itself designate a property. Hence the

property in virtue of which an object satisfies the predicate 'is in pain' ... is not the generic property of being in pain. There is no such property. ... The idea, rather, is that 'is in pain' applies to creatures that are similar in certain salient respects: similar enough to merit application of the pain predicate. These similarities stem from the creatures' possession of certain properties. The properties need not be the same in every case: creatures are neither

identical nor exactly similar in those respects in virtue of which they satisfy the predicate ‘is in pain.’ (Heil 1999, 201).

‘Higher-order’ predicates such as functional predicates, thus, describe objects more or less abstractly, but “to describe an object abstractly is not to ascribe some further property to the object” (Heil 1999, 202).

On this conception, multiple realizability is not, as it is standardly thought to be, a relation among properties. It is simply a fancy name for the familiar phenomenon of predicates applying to objects in virtue of the possession by those objects of distinct, although pertinently similar, properties. (Heil 1999, 203)

Thus, the predicate ‘is in pain’ can still be employed in theoretically fruitful generalizations, even generalizations that would remain invisible at lower levels of abstraction, and can even be projectible in a lawlike way where this *ceteris paribus* lawlikeness is grounded in the similarity of causal powers picked out by the predicate.

This, then, is the determinable/determinate account of multiple realizability (or at least one version thereof). First, let us ask how this account should be applied to qualia, construed as we have done not as functional properties but as phenomenal ones. Then, let us try to see what light, if any, this sheds on the natural kind status of qualia.

We shall take, as is traditional, pain as our central example. And, as it is also traditional to remark, pain is pretty clearly multiply realizable, appearing as it does in all sorts of different actual and merely possible species with very different neurological make-ups. If we are talking of pain as if it were a putative natural kind ‘stuff,’ then we can say that its micro-constitution is apparently very different in various of its different instantiations. What then, is the status of *pain* vis-à-vis its various manifestations? What is the metaphysical status of pain as a ‘higher-level property’? On the determinable/determinate view, pain is *not*, in fact, an additional property, but is (nothing more than) the name we give to the pattern of *similarity* exhibited by all the various objects or states to which we apply the term. Put another way, the world contains many determinate properties—each of whose instances are *identical* in a certain respect with each other—and clusters of these determinate properties evince a ‘family resemblance’—a similarity but *not* an identity—in virtue of which we sometimes group them together under the umbrella of a single predicate. The world, thus, contains normal human pain, which is perhaps a particular neural type; it also contains hydrocephalic human pain, which might be another neural type; and it contains normal octopod pain, another, perhaps, non-myelinated neural type; and so on more or less indefinitely. What all these *different* properties

have in common is not some sort of tokening of a further ‘higher-level’ property—which is, as Heil recognises, a metaphysically rather slippery notion in any case—but a kind of abstract similarity. In what respect, then, are the various different incarnations of pain similar? One way in which they may all be similar, of course, is that they fit a certain sort of functional profile, but *qua* qualia as it were, what we require is a sort of *phenomenal* or *qualitative* similarity. All the different properties enumerated above deserve to be grouped under the predicate “pain” insofar as they all *feel painful* (as opposed to ticklish, or tasty, or red, or sticky, and so on). Further, part of *feeling painful*, plausibly, is that the sensation is, *ceteris paribus*, a distasteful one—one whose repetition we seek to avoid. Something simply could not be a *pain* sensation, for some species, if it was a sensation that members of that species routinely found enjoyable and sought out. To say this is not, incidentally, to turn to specifying qualia by their functional role—rather, it is to attempt to say a little more about the nature of the qualitative ‘feel’ that is involved in painfulness. It does, however, begin to connect the phenomenality of qualia with certain causal powers that those properties might have. Put it this way: though qualia may not be exhaustively specifiable purely in virtue of their causal relations (contra functionalism), it is still open to us to hold that that intrinsic nature of qualia—the painfulness of pain, for example—contributes to their having the causal powers they do.

This qualitative similarity may not be a *property* properly speaking, on the determinable/determinate view, but it is nevertheless a perfectly ‘real pattern.’ And it is a pattern that can support scientific explanations and generalizations. For any species that can feel pain and whose members are capable of learning, for example, perhaps painful stimuli can always be used as a way of conditioning behaviour (all things being equal). This is, if true, a perfectly good lawlike generalization, and it is furthermore one that would have remained invisible—indeed impossible to formulate—in terms of any one of the determinate *properties* that all go by the name of “pain.”

In what way, if any, does this help the cause of making qualia ‘scientifically respectable’? It is helpful, first of all, insofar as it tends towards the removal of the aura of metaphysical mystery surrounding apparently ‘higher-level’ mental ‘properties,’ especially their apparent causal impotence over and above the causal powers of their realizers. Pain is not an epiphenomenal interloper, on this view, but a perfectly real and scientifically significant pattern of similarity between determinate properties that have causal powers. Furthermore, the *way* in which these properties are similar accounts for the way in which their causal powers resemble each other—to put it very roughly, pain, in all its different species-specific manifestations tends to bring about avoidance behaviour *because it hurts*.

I said at the outset that something is a natural kind, roughly if it is a *scientifically respectable* kind: if it is a kind whose members all have some kind of essential similarity, and form part of a natural scientific domain governed by laws and regularities that support, for example, inference to counterfactual cases, and so on. On this accounting, it is, at least, not clear that qualia are *not* natural kinds: pain, for example, is a group of properties—each of which has the usual credentials as a member in good standing of the physical world’s causal economy—and these properties stand in a certain kind of similarity relation to each other. Furthermore, this similarity underpins and explains a set of perfectly legitimate scientific explanations and predictions, using lawlike regularities that are invisible if the determinate properties are taken as individuals outside of the group. These lawlike regularities may not be grounded in shared *properties* that all the members of the group possess, but to recognise this is perhaps merely to recognise that an older way of talking has become outmoded. After all, the determinable/determinate model of realization, if it holds at all, holds for a very wide range of phenomena well beyond the mental, such as colour science or geology, for example. As Fodor recognised in 1974, if we are to deny scientific legitimacy to one of the autonomous ‘special sciences,’ it seems we must deny it to all.

Finally, all of this also suggests what is, to me, an intriguing approach to the naturalization of qualia. Suppose we want to explain the similarity between structure-specific realizations of a certain quale, such as, once again, pain. On the natural kind model, we might do this by empirically exploring the microstructure of the various different instantiations: if pain is generally qualitatively unpleasant, and thus a state that tends to lead to aversion behaviours of various sorts, what is it that the different kinds of ‘stuff’ that realize pain have in common such that they all have these causal powers? This is not, of course, to say that all pain realizations must have the *same* underlying nature—that is, this is not to *deny* the multiple realizability of pain. Martian pain and human pain might—indeed, in virtue of their very different realizations, probably would—feel significantly different, and have significantly different causal roles in the respective Martian and human mental economies. But nevertheless, their feel and causal role will be sufficiently *similar* (otherwise, we should simply say that Martians do not feel pain). The naturalization of qualia, then, will consist in part in explaining this similarity.

References:

Block, Ned (1980). "What is Functionalism?"

Capitan, W.H., and D.D. Merrill, eds. (1967). *Art, Mind, and Religion*, Pittsburgh: University of Pittsburgh Press.

Heil, John (1999). "Multiple Realizability," *American Philosophical Quarterly* 36. 189–208.

Kim, Jaegwon (1992). "Multiple Realizability and the Metaphysics of Reduction," *Philosophy and Phenomenological Research* 52. 1–26.

Putnam, Hilary (1967/1975). "Psychological Predicates," in Capitan and Merrill 1967, 37–48.
Reprinted as "The Nature of Mental States" in Putnam 1975, 429–40.

Putnam, Hilary (1975). *Mind, Language and Reality: Philosophical Papers* vol. 2, Cambridge: Cambridge University Press.

Richardson, Robert C. (1979). "Functionalism and Reductionism," *Philosophy of Science* 46. 533–558.

Robb, David (1997). "The Properties of Mental Causation," *The Philosophical Quarterly* 47. 178–194.