WILLIAM JAWORSKI

# MULTIPLE-REALIZABILITY, EXPLANATION AND THE DISJUNCTIVE MOVE

(Received in revised form 4 July 2001)

ABSTRACT. The multiple-realizability argument has been the mainstay of anti-reductionist consensus in philosophy of mind for the past thirty years. Reductionist opposition to it has sometimes taken the form of the *Disjunctive Move*: If mental types are multiply-realizable, they are not coextensive with physical types; they might nevertheless be coextensive with disjunctions of physical types, and those disjunctions could still underwrite psychophysical reduction. Among anti-reductionists, confidence is high that the Disjunctive Move fails; arguments to this effect, however, often leave something to be desired. I raise difficulties for one anti-reductionist response to the Disjunctive Move, the *Explanatory Response*.

## 1. PSYCHOPHYSICAL REDUCTION AND THE MULTIPLE-REALIZABILITY ARGUMENT

The Explanatory Response is an anti-reductionist argument originally suggested by Putnam (1975, 1981), and certain biologists and philosophers of biology.[1] The argument has been advanced in more developed forms by Kitcher (1984) contra the reduction of classical genetics to molecular biology, and by Pereboom and Kornblith (1991) contra psychophysical reduction. It is opposition to the latter that interests us here.

Psychophysical reduction is the thesis that psychology is reducible to physical theory. 'Psychology' here refers to the conjunction of theories or conceptual frameworks that use mentalistic terminology in the construction of true explanations of animal behavior.[2] 'Physical theory', by contrast, refers to the conjunction of true theories or frameworks that do not use such terminology; these include theories of modern physics, chemistry, and molecular biology, as well as theories of nervous functioning, embryonic development, immunology, and the like. On the standard model of reduction (Nagel, 1961), the fact that the descriptive vocabularies of

*Philosophical Studies* **108:** 289–308, 2002.

psychology and physical theory are at least partially disjoint means a reduction of one to the other must involve "bridge principles" linking their respective terms. Although Nagel did not require bridge principles to be anything more than conditionals, it is commonly assumed they must take the form of either identity statements or biconditionals that are true in at least every nomologically possible world. If psychophysical reduction is true, there is for every mental term 'M', a physical term 'P' such that at least (1) is true and possibly (2):

(1) $\Box$ (x). Mx $\leftrightarrow$ Px
(2) M = P

where strength of the modal operator in (1) is at least nomological.

Given that (1) is entailed by (2), an argument establishing the denial of the former would establish also the denial of the latter, and thereby sweep psychophysical reduction under the rug. This is precisely what the multiple-realizability argument (MRA) purports to do:

> ($MR_1$) Mental types are multiply-realizable relative to physical types;
> ($MR_2$) If mental types are multiply-realizable relative to physical types, mental and physical types are not nomologically coextensive;
>
> ---
>
> ($MR_3$) Mental and physical types are not nomologically coextensive.

A mental type, M, is multiply-realizable relative to physical types just in case, possibly, M tokens are realized by physical tokens belonging to more than one type, where the strength of the modal operator is nomological.[3]

Whatever it means for one token to realize another, the crucial point is that a realizer is sufficient for what it realizes.[4] Suppose, for instance, that Alexander's pain is realized by his c-fibers firing. If pain is multiply-realizable, as per $MR_1$, there is a nomologically possible world in which pain is realized by something other than c-fiber firing. It seems possible, for instance, that there might be an alien species of pain-capable organisms whose pains are realized by the firing of q-fibers ($\neq$ c-fibers). Realizers are sufficient for the tokens they realize, so c-fiber firing and q-fiber firing are each

sufficient for pain. In that case, however, neither c- nor q-fiber firing by itself is necessary, so the following conditionals are false:

(3) $\Box$ (x). x is in pain $\rightarrow$ x has c-fiber firing;
(4) $\Box$ (x). x is in pain $\rightarrow$ x has q-fiber firing.

The falsity of (3) rules out a bridge principle that would link pain with c-fiber firing in the manner of (1) or (2), and the falsity of (4) does the same with respect to q-fiber firing. Since psychology reduces to physical theory only if mental types are nomologically coextensive with physical types, psychophysical reduction is out of the question.

## 2. THE DISJUNCTIVE MOVE

The bulk of reductionist responses to the MRA have focused on its second premise.[5] One such response is the *Disjunctive Move*.[6] If mental types are multiply-realizable relative to physical types, then at best each mental type will be correlated with a disjunction of physical ones. According to exponents of the MRA, this rules out psychophysical reduction. What they assume is that such disjunctions cannot themselves service reduction. But, a reductionist might urge, why assume that? If pain is realizable by c-fiber firing and q-fiber firing, the following biconditionals will indeed be false:

(5) $\Box$ (x). x is in pain $\leftrightarrow$ x has c-fiber firing;
(6) $\Box$ (x). x is in pain $\leftrightarrow$ x has q-fiber firing.

But the following biconditional might still be true:

(7) $\Box$ (x). x is in pain $\leftrightarrow$ .x has c-fiber firing $\vee$ x has q-fiber firing.

Might not (7) function as a bridge principle? If so, it appears that multiple-realizability may be compatible with reduction after all. In more general terms, if mental types are multiply-realizable relative to physical types, there is a true biconditional of the following form for each:

(8) $\Box$ (x). $Mx \leftrightarrow .P_1x \vee P_2x \vee \ldots \vee P_nx$,

where M is the mental type, and each $P_i$, a physical type whose tokens realize an M-token in some nomologically possible world. Advocates of the Disjunctive Move propose that biconditionals like

(8) be used as bridge principles for psychophysical reduction, and challenge anti-reductionists to explain why they can't.[7]

## 3. THE ANTI-REDUCTIONIST TARGET

There are several ways anti-reductionists might respond to the Disjunctive Move. One way tries to nip it in the bud by denying either that mental tokens must be realized, or that they must be realized physically in all nomologically possible worlds. But this response does not interest us here. For the purposes of this paper, we will assume the following:

(i) Necessarily$_M$, if there are mental tokens, they are realized;
(ii) Necessarily$_N$, if there are mental tokens, they are realized physically.

The operators 'Necessarily$_M$' and 'Necessarily$_N$' cover metaphysically and nomologically possible worlds, respectively. Assumption (i) seems common enough among friends of the MRA.[8] Assumption (ii) is more controversial, though it seems a reasonable expectation of non-reductive physicalists. The idea, I take it, is that the laws of nature impose constraints on the sorts of things that can realize mental types,[9] with the result that worlds in which those laws obtain are worlds in which mental types are realized physically or not at all.

Assumption (i) implies that mental tokens must (metaphysically) be realized if they exist, and (ii) implies they must (nomologically) be realized physically. $MR_1$, on the other hand, implies they must be realized by tokens belonging to more physical types than one. As a result, Anti-reductionists who endorse (i), (ii), and $MR_1$ are committed to the existence of biconditionals like (8). In response to the Disjunctive Move, then, friends of the MRA who endorse (i) and (ii) must target the claim that those biconditionals can function as bridge principles.[10] One argument to this effect is the *Explanatory Response*.

## 4. THE EXPLANATORY RESPONSE

To understand the Explanatory Response, we should first say a word about the strategy behind it. In an influential paper, Fodor (1974) articulated what he took to be a key component implicit in the reductionist program: in cases of genuine reduction, bridge principles must connect natural kinds, in particular, kinds postulated by the reduced theory with kinds postulated by the reducing theory. Since natural kind predicates are just the non-logico-mathematical predicates featured in scientific laws, this is simply to say that in genuine reductions, bridge principles must be genuine laws. For anti-reductionists who accept this idea (and our focus here is on them), the strategy is to argue that biconditionals like (8) fail to be laws, or equivalently, that the disjunctive predicates they feature ("R-disjunctions" henceforth) fail to express natural kinds. The result, if their arguments succeed, is that biconditionals like (8) cannot underwrite psychophysical reduction. The Explanatory Response, in particular, claims such biconditionals cannot be laws because laws are explanatory, and generalizations featuring R-disjunctions *ipso facto* are not. Pereboom and Kornblith ("P&K" henceforth) state the argument thus:

> If kinds in psychology are multiply realizable ... purported bridge principles relating psychological to neurophysiological kinds will involve open-ended disjunctions ... Why are such disjunctions not natural kinds? ... because they cannot appear in laws. They cannot appear in laws because "laws" involving such disjunctions are not explanatory ... they do not meet our interests in explanation.[11]

Much of the ensuing discussion focuses on P&K's formulation of the argument.

The traditional covering-law model of explanation took the latter to consist in logical derivation. Laws explain facts (general or particular) in the sense that the latter are logically derivable from the former together with statements describing initial conditions. Statements of the form "$\varphi a$ because $p$" or "$\forall x \varphi x$ because $p$" (where $p$ is some statement, and $\varphi a$ and $\forall x \varphi x$ are statements expressing particular and general facts, respectively) are, therefore, elliptical means of expressing derivations from lawlike generalizations. Friends of the Explanatory Response, however, cannot claim that logical

derivability is sufficient for explanation. Given a biconditional like (8), $\varphi a$ could be derived from an R-disjunctive generalization like

(9) $\Box$ (x). $P_1x \vee P_2x \vee \ldots \vee P_nx. \rightarrow \varphi x$,

just as easily as it could from a generalization featuring the corresponding mental predicate:

(10) $\Box$ (x). $Mx \rightarrow \varphi x$.

It would be awkward for anti-reductionists if both sorts of derivations, or neither, turned out to be explanatory (awkward if both, because that would sanction reduction; awkward if neither, because in that case the notion of *psychological explanation* would be incoherent). Hence, friends of the Explanatory Response must claim there is a qualitative difference between derivations from R-disjunctive generalizations like (9), and derivations from psychological ones like (10), a difference that renders the latter, but not the former explanatory. If they have no qualms about the necessity of derivation for explanation (which we will assume here), Explanatory Response advocates must claim that explanation consists in derivation *plus* something else.

It is, of course, widely acknowledged that not all derivations from laws are explanatory. There is still no consensus, however, on a principled way of distinguishing ones that are from ones that are not – and this, in large part, because there is still no consensus on just what explanation in general is supposed to consist in, or even whether it is reasonable to expect a general account of explanation that abstracts from the methods and practices of concrete disciplines.[12] That aside, there is still a good deal of plausibility in the claim that whatever else derivations must do to qualify as explanations, they must pick out factors that are *relevant*. Putnam illustrated this using the celebrated example of a square peg and round hole:

> . . . [Suppose we have] a board in which there are two holes, a square hole one inch across and a round hole one inch in diameter, and a square peg, a fraction less than one inch across. The fact to be explained is this: The peg goes through the square hole, and it does not go through the round hole . . . One explanation is that the peg is approximately rigid under transportation and the board is approximately rigid. The peg goes through the hole that is large enough [to take its cross-section] and not through the hole that is too small . . . Suppose, however, we describe the board as a cloud of elementary particles . . . and imagine ourselves given the position

and velocity at some arbitrary time $t_0$ of each one. We then describe the peg in a similar way. (Say the board is "cloud B" and the peg is "cloud A".) Suppose we describe the round hole as "region 1" and the square hole as "region 2".[13]

If the square is in fact composed of fundamental-physical items, there should be a derivation (however unique) from the laws of fundamental physics that proves Cloud A will pass through Region 2 but not through Region 1. But should we consider such a derivation explanatory? Clearly not, says Putnam. Explanations are supposed to reveal *relevant* relations, in this case geometrical ones; a derivation cast in fundamental-physical terms, however, reveals nothing about the geometry of the situation, and is therefore not an explanation. Kitcher (1984) makes a similar point regarding classical genetics (though regrettably the example is not as lucid as Putnam's):

> . . . I claim that exhibiting derivations of the transmission laws from principles of molecular biology and bridge principles would not explain the laws, and, therefore, would not fulfill the major goal of reduction.
>
> As an illustration, I shall use the envisaged amended version of Mendel's second law. Why do genes on nonhomologous chromosomes assort independently? Cytology provides . . . a perfectly satisfactory explanation of why our envisaged law is true to the extent that it is . . . To emphasize the adequacy of the explanation is not to deny that it could be extended in certain ways . . . However, appeal to molecular biology would not deepen our understanding of the transmission law. Imagine a successful derivation of the law from principles of chemistry and a bridge principle . . . In charting the details of the molecular rearrangements the derivation would only blur the outline of a simple cytological story, adding a welter of irrelevant detail.[14]

According to Kitcher, the derivation of a (suitably corrected) law of classical genetics from molecular biology would, at best, add nothing to the understanding we gain from a cytological derivation, and would, at worst, obfuscate the understanding we might gain from the latter. It would, in short, fail to be relevant in a way that explained the genetic law.

P&K claim the case of psychology is no different from the foregoing ones. Even if psychological facts could be derived from R-disjunctive generalizations, they say, those derivations would not be explanatory, for they would fail to reveal relevant relations. In support of this, they cite the following example:

When Mary walks down the street to buy an ice-cream cone, we explain her behavior by appealing to the content of her beliefs, and desires: she wanted an ice-cream cone and she believed one could be purchased down the street. Replacing this explanation by one which contains an open-ended disjunction of physical predicates – if Mary is in state $P_1$ or $P_2$ or $P_3$, etc., she will move with trajectory $T_1$ – indeed leaves our interests in explanation unsatisfied.[15]

The example is meant to expose the intuition that derivations couched in R-disjunctive terms do not satisfy our interests in ordinary psychological explanation. It is fair to say the example succeeds; it seems *obvious* that an R-disjunctive derivation of Mary's walking is not explanatory the way a psychological derivation is. The latter isolates features of the situation that are relevant to our interests in explaining normal human behavior; the former does not. This is just to say, however, that appeals to psychological generalizations are explanatory, while appeals to the R-disjunctive ones are not. As a result, the argument claims, R-disjunctive generalizations cannot be laws.

To summarize, then, the premises of the Explanatory Response are as follows:

($ER_1$) Laws are explanatory;
($ER_2$) A generalization is explanatory only if it can factor into relevant derivations;
($ER_3$) R-disjunctive generalizations cannot factor into relevant derivations.

If explanation consists in relevant derivation, and R-disjunctive generalizations cannot factor into relevant derivations, then they cannot be explanatory, and hence cannot be laws. But if they cannot be laws, *a fortiori* they cannot be bridge laws. Consequently, biconditionals like (8) cannot function as bridge principles for psychophysical reduction, and the Disjunctive Move fails.

## 5. PROBLEMS WITH THE EXPLANATORY RESPONSE

For our purposes, we will take $ER_1$ as a plausible constraint on the concept of a law. Consider, then, $ER_2$ and $ER_3$. Reductionist objections to $ER_2$ will try to show that generalizations need not factor into relevant derivations to be explanatory. In this connection, one is puzzled to find P&K striving to rebut an argument

of a different sort.[16] Science is ultimately concerned with delivering a true account of how the world is, the argument claims, so the explanations it generates must be objective. The relevance of a derivation, however, depends on whatever explanatory interests we have in a given situation, and those interests, far from being objective, vary greatly from one situation to the next. Consequently, relevance cannot be the whole story about explanation.

P&K respond to the argument by attacking the premise that our explanatory interests are not objective. Not only can interests be objective, they say, but the success of a scientific research program actually provides evidence of the objectivity of the interests that generated it. The latter claim is dubious, or at least ambiguous. One wants to know, for instance, whether research programs generated by interests in financial gain, say, are guaranteed failure, or whether programs generated purely by interests in gaining scientific knowledge are guaranteed success. What is most puzzling, however, is why P&K undertake to rebut the argument in the first place. After all, what it attacks is not the necessity of relevance for explanation (which is all the Explanatory Response needs), but only its sufficiency. The argument merely claims that relevance cannot be the whole of what makes a derivation explanatory. Friends of the Explanatory Response can happily agree, provided they insist relevance is at least necessary.

P&K's worries stem, perhaps, from a conflation of the foregoing argument and another they mention:

> If only we were capable of taking in more information at once, the reductionist might say, we wouldn't have any trouble regarding open-ended disjunctive "laws" as genuine laws. The fact that we fail to find laws satisfying when they contain open-ended disjunctions may simply show a failing on our part, rather than on the part of the laws themselves.[17]

Interpreted one way, this argument would challenge the necessity of relevance for explanation, for it would claim that our concern with relevant derivation stems from cognitive limitations peculiar to us. If our intellectual powers were not so blunt, it would claim, the "relevance" of derivations in P&K's sense would not be an issue; relevance is not, therefore, a necessary feature of explanation at all. Yet the conclusion of this argument seems so implausible as to militate against the interpretation on which it is based. Could anyone

seriously think that genuine explanations might be quite *irrelevant*? Or that appeals to a given generalization might always be irrelevant and yet occasionally explanatory? Looking at the matter this way invites the suggestion that we accept $ER_2$ as a reasonable constraint on the notion of explanation, and find another interpretation for the argument.

A more plausible interpretation has the argument claiming not that we would find relevance unnecessary for explanation if we were cognitively better off, but only that we would find relevance in different things, e.g. in derivations from R-disjunctive generalizations. Interpreted this way, the argument challenges not $ER_2$, but $ER_3$, the premise that R-disjunctions can never factor into relevant derivations. This, I think, is where the real problems with the Explanatory Response lie.

At least two problems center on $ER_3$. The first concerns the scope of the argument. The Explanatory Response purports to show that R-disjunctive generalizations cannot be laws since derivations cast in terms of them fail to satisfy our explanatory interests. As we have seen, Explanatory Response advocates like P&K give examples to support their claim, and there can be little doubt that those examples make a strong intuitive bid. It seems obvious that an explanation of Mary's walking couched in psychological terms satisfies our interests in a way a derivation in R-disjunctive terms does not. But why exactly is that? What interests do we have in such contexts? And why do R-disjunctions fail to address them?

Regarding the first question, it is plausible to suppose that in ordinary contexts in which we try to make sense of one another's behavior, we are interested in obtaining an immediate, intuitive grasp of other agents' motivations, perhaps by acquiring information that allows us to project ourselves imaginatively into their circumstances.[18] But why do derivations involving R-disjunctions fail to provide that? P&K cite three features of R-disjunctions that are significant: they are disjunctive; they are open-ended, and they are physicalistic. Here 'physicalistic' means simply that R-disjunctions feature predicates of physical theory and none of psychology, and 'open-ended' means that each R-disjunction involves, or might involve, an indefinite, perhaps infinite, number of disjuncts. Which of these features (or which combination of them)

is responsible for the failure of R-disjunctive derivations to satisfy our interests? On this point, an historical observation is in order.

The intuition that physicalistic considerations (in the foregoing sense of "physicalistic") are irrelevant vis-à-vis our interests in psychological explanation has a venerable tradition dating back at least to Plato. Consider the following passage from the *Phaedo* in which Plato's Socrates recounts his experience reading Anaxagoras:

> . . . My wondrous hopes were swept away, my friend, when I proceeded to read, and saw the man neither appealing to mind, nor citing any of the causes responsible for the ordering of things, but instead citing air, and aether, and water, and many other absurdities as causes. To me it seemed exactly the same as someone saying that Socrates does everything he does with mind, and then in undertaking to state the causes of each thing I do were to say that I am sitting here now because, first, my body is composed of bones and sinews, and the bones are hard and have joints separating them, while the sinews for their part contract and relax, and cover the bones along with the flesh and skin that contains them, and that because the bones move freely in their joints, the contracting and relaxing of the sinews somehow enables me to bend my limbs now, and this is the cause of my sitting here in a bent position . . . But to call such things causes is most absurd. If someone were to say that without having bones, and sinews, and such, I would not be able to do what I believe best, that would be true. But to say that I do what I do *because* of these, and therein act with mind, but not on account of choosing what I believe best – *that* would be an extremely careless way of speaking (98c–99b).[19]

Socrates' complaint stems from the fact that Anaxagoras' accounts of human behavior, which are couched in exclusively physical terms, do not address his explanatory interests. What interests Socrates are the *reasons* for people's actions, not the physical mechanisms by which they move, which are all that accounts like Anaxagoras' address. What is significant is that Socrates faults Anaxagoras' so-called "explanations" not on account of their being disjunctive or open-ended, but simply on account of their being physicalistic.

We encounter a similar point in the following:

(11) Mary is walking down the street because neurons $n_1, \ldots, n_j$ are in states $s_1, \ldots, s_j$ and muscles $m_1, \ldots, m_n$ are contracting at rates $r_1, \ldots, r_j$.

Statement (11) seems unsatisfying as an explanation of Mary's walking, and yet is neither disjunctive nor open-ended. On the other

hand, the following statement *is* disjunctive and open-ended, but not obviously unsatisfying as an explanation:

(12) Mary is walking down the street because she wants an ice cream cone, and believes she can purchase one down the street, or perhaps because she believes they are spying on her, and wants them to think she enjoys ice cream, or perhaps because she wants to see Pedro, the ice cream man, and believes walking to get a cone an excellent pretense, or perhaps . . .

The point is not that (12) certainly does provide an explanation of Mary's walking, only that its failure is not obvious – at least not as obvious as that of (11).

Finally, anti-reductionists have sometimes claimed that statements like (11) fail to be explanatory because they include too many details.[20] Waters (1990) refers to this as the "Gory Details Argument". What is significant about the argument is that it too suggests that physicality is the feature of R-disjunctive generalizations that renders derivations from them unexplanatory. The morass of physical details we meet in the likes of (11) is due not to the fact that the putative explanantia are disjunctive or open-ended, but to the fact that the physical categories in which they are framed are finely grained, and hence generate accounts too tortuous to follow with any ease or facility.

What the foregoing considerations suggest is that physicality may be the crucial feature of R-disjunctions that renders derivations from them unsatisfying as explanations. Being open-ended or disjunctive might make for unwieldy derivations, but physicality, it seems, is sufficient. In that case, however, there is reason for being suspicious of the Explanatory Response. If physicality is sufficient to render R-disjunctive derivations unexplanatory, it appears that the Explanatory Response generalizes to *all* derivations couched in physical terms. But then the argument ends up proving too much, for it concludes that physical generalizations *in general* cannot be explanatory, which is absurd. Friends of the Explanatory Response could reply by weakening $ER_3$, as in:

($ER'_3$) R-disjunctive generalizations cannot factor into relevant derivations *of particular psychological facts*.

They could claim, in other words, that their intention was never to prove that R-disjunctive generalizations could not be explanatory at all, but only that they could not be used to explain particular psychological facts. But then the argument would prove too little, for defeating the Disjunctive Move requires showing that generalizations featuring R-disjunctions not be laws at all, or equivalently, that R-disjunctions not express natural kinds of any sort. If the Explanatory Response is relativized to derivations of particular psychological facts as in $ER_3'$, the argument shows only that R-disjunctive generalizations cannot be laws of psychology, or that R-disjunctions cannot express psychological kinds. But that hardly touches the Disjunctive Move, for the possibility lingers that R-disjunctions might express kinds of some other sort, and that is all friends of the Disjunctive Move need to claim biconditionals like (8) are genuine bridge laws. Friends of the Explanatory Response have no choice, it seems, but to fight hard against the claim that physicality is the crucial feature rendering R-disjunctive derivations irrelevant. But this promises to be an uphill battle, for given the likes of the Gory Details Argument, it would appear to require a re-regimentation of Anti-reductionist forces.

The second problem with $ER_3$ was suggested by an argument mentioned earlier. If we were better off cognitively, the argument claimed, we would find relevance in different derivations, e.g. ones involving R-disjunctions. This suggests a general strategy for dealing with the Explanatory Response, one that trades on the following points: (i) in different contexts, there may be different criteria of relevance, and (ii) examples like P&K's Mary case concern the criteria of relevance for one type of context only. Consider one way of applying this strategy.

In the context of ordinary psychological explanation, there can be little doubt that the Explanatory Response has a lot going for it: in those contexts, our interests are clearly not satisfied by R-disjunctive derivations. But "ordinary" contexts are not the only ones that concern us here. After all, reductionists never claimed (or ought never to have claimed) that R-disjunctive derivations would supplant psychological ones in our quotidian dealings. Such derivations were never meant to provide explanations of human behavior in pedestrian contexts at all, but rather in the highly specialized

context of intertheoretic reduction. In reductive contexts, however, the interests we have might be different from those we have in more pedestrian contexts. This is significant, for *if our interests in reductive contexts differ from those in pedestrian ones, reductive explanations will not have to meet the same criteria of relevance*. As a result, the strong intuitions we meet in the likes of P&K's Mary example will be neither here nor there, and derivations from R-disjunctive generalizations will have a shot at relevance after all.

Recall our conjectured criterion of relevance for explaining particular psychological facts in a pedestrian context: explanations should provide an immediate, intuitive grasp of a given psychological fact. Clearly, there are reductive contexts in which we expect something similar, e.g. we often expect microphysical explanations to provide an intuitive grasp of certain macrophysical facts. The question, however, is whether this is *always* what we expect in reductive contexts. Might there not be reductive contexts in which an intuitive grasp is something we are willing to do without? Consider one possible case.

Suppose an advanced alien species has sent emissaries to earth bearing a complete physical theory as a gift – call it 'T*'. We discover that from T* we can derive and successfully predict all the phenomena our current physical theories can explain, plus several besides, and that T* does so within a framework beholden to fewer basic explanatory principles, and with fewer ontological commitments than our current theories. To what extent are we interested in T* giving us an "intuitive grasp" of the physical facts? (Ask yourself to what extent we expect this of quantum theory.) The point isn't that we will certainly not be interested in T* providing us with such a grasp, only that we *need* not be. In such a case, we might be content knowing simply *that* the theory works without knowing in an immediate, intuitive way exactly how; our interests, in other words, might amount to a demand for nothing more than simple derivation – that might be our sole criterion of relevance.

The question, then, is whether things might not be similar in the case of psychology. Would it be enough to know that from the laws of physical theory we could derive and successfully predict all human behavior? Would our interests in that particular reductive context be satisfied? Anti-reductionists will want to claim they

would not, that in the case of psychophysical reduction, we expect (or should expect) exactly the same sort of intuitive insight gained from, say, microphysical explanations of disease or principles of thermodynamics.[21] Reductionists should respond that such expectations are unreasonable. No one would seriously doubt the reducibility of chemistry to fundamental physics, yet given the complexities involved, it would be altogether unreasonable to demand that derivations from the principles of quantum mechanics should elucidate laws of chemistry the way anti-reductionists would have derivations from physiological or biochemical principles elucidate laws of psychology. In reductions of great complexity (as psychophysical reduction promises to be), it is enough to know that the derivations work, even if we can't understand in an intuitive way exactly how. If reductionists carry the day here, it looks like the Explanatory Response fails, for in that case, $ER_3$ proves false: if straightforward derivation satisfies our interests in the context of psychophysical reduction, R-disjunctive generalizations *can* factor into relevant derivations, and can, therefore, be laws.

To summarize: The Explanatory Response claims that laws must be able to factor into relevant derivations, and that means derivations that suit our interests. But those interests may differ from one context to the next. Relative to the interests we have in pedestrian contexts, the Explanatory Response stands in good stead. In those contexts we are interested in information that allows us to grasp in an immediate, intuitive way other agents' motivations, and there can be little doubt that R-disjunctive derivations fail by those lights. The problem is that those are not the only lights by which relevance is to be reckoned. Reduction provides contexts in which an immediate, intuitive grasp of a given domain of facts may not be what interests us. There may be reductive contexts in which all we care about is simple derivation, in which the latter provides as much in the way of explanation as we desire. In such contexts, R-disjunctive generalizations may turn out to be explanatory after all, and hence nothing prevents them from being laws.

To resist the foregoing argument, anti-reductionists could deny there are reductive contexts in which our explanatory interests differ from those in pedestrian contexts. This would require them, of course, to insist on the reasonableness of expecting all reductions

to supply roughly the same level of intuitive insight we achieve in our ordinary dealings with one another. But that seems a hard sell given the highly specialized nature of intertheoretic reduction. Short of that, they could try denying that psychological generalizations are derivable from R-disjunctive ones. But given the assumptions mentioned in §3, there must be R-disjunctive biconditionals like (8), and that means psychological generalizations like

(13) $\Box$ (x). $Mx \rightarrow M^*x$,

for mental types M and M*, will be derivable from generalizations featuring the corresponding R-disjunctions as antecedents and consequents:

(14) $\Box$ (x). $P_1x \vee P_2x \vee \ldots \vee P_nx. \rightarrow .P_1^*x \vee P_2^*x \vee \ldots \vee P_n^*x$,

where the $P_i$ and $P_i^*$ are those physical types whose tokens realize M tokens and M* tokens, respectively, in some nomologically possible world. Anti-reductionists could object here that most likely there is not a law of physical theory like (14), or equivalently, that most likely '$P_1 \vee P_2 \vee \ldots \vee P_n$' and '$P_1^* \vee P_2^* \vee \ldots \vee P_n^*$' do not express physical natural kinds.[22] Reductionists could respond, of course, by *adding* such laws to physical theory. If successful, the Explanatory Response would provide grounds against this sort of addition, but of course, the success of the Explanatory Response is precisely what is at issue.

Anti-reductionists could also jettison one of the aforementioned assumptions: the claim that necessarily$_M$, mental tokens are realized, if they exist, or the claim that necessarily$_N$, if they exist, they are realized physically. But they would do so at the risk of jettisoning physicalism as well, and that might be undesirable. Anti-reductionists could also tamper with the notions of realization and reduction, e.g. by denying that realizers are sufficient for what they realize, or by requiring that bridge principles take the form of identities. The former option, however, is clearly a non-starter, for it undermines the MRA's second premise. The latter option, on the other hand, has something going for it, for there are good reasons for thinking that reduction does in fact require identities.[23] The only problem is that the success of derivations via biconditionals like (8) itself provides grounds for "upgrading" those biconditionals to identities.[24]

By contrast, the only commitments reductionists shoulder here are the claims that (a) we need not expect all reductions to conform to the same standards of explanatory adequacy we have in ordinary psychological contexts, and (b) there might be laws of physical theory along the lines of (14). Given the options, the likely conclusion is that the Explanatory Response fails. Anti-reductionists would do better on a different tack.

## NOTES

[1] The outlines of the argument are apparent, for instance, in Hull (1974), and Mayr (1969) (selections of which are quoted in Hull).

[2] Mental terms, we might say, are ones connoting either intentionality or consciousness. The former terms often include intentional verbs, those typically taken to mark propositional attitudes (e.g. 'thinks', 'believes', 'knows', 'wishes', 'fears', 'hopes', 'remembers', 'desires', 'perceives', 'wants', 'regrets'), and often generate nonextensional contexts, where the latter are contexts in which the usual rules for substitution of equivalent expressions and for existential generalization fail to apply. Terms connoting consciousness, on the other hand, typically include those describing sensory states, both those applicable to the perceiver (e.g. "is in pain", "hears something like running water"), and those applicable to the object perceived (e.g. "feels sharp", "tastes like chicken").

[3] Anti-reductionists might endorse a weaker claim here by having the modal operator cover metaphysically or logically possible worlds. But the claim that it is possible, the laws of nature being what they are, for there to be entities (e.g. artifacts, non-human animals) that are psychologically indistinguishable from humans, but nevertheless physically very different seems common enough among friends of the MRA. See, for instance, Putnam (1967), p. 228, Putnam (1975), pp. 135–136, Block and Fodor (1972), p. 238, and Fodor (1974), p. 125.

[4] Several analyses of the realization relation have appeared in the literature. See Putnam (1970), Block (1990), Kim (1992), and LePore and Loewer (1989) for examples.

[5] Reductionist responses to the first premise include attacks on the intuition that cases like the pain-capable Martian are possible (Kim 1972, p. 235), postulation of "broader" physical types (Kim, *op. cit.*), and attacks on the claim that intuitions about mental types can function evidentially in this arena at all. The main challenge in the latter connection stems from the extension of scientific essentialism to psychological types; see Bealer (1994).

[6] Kim (1978, 1984, 1990) has been the most outspoken advocate of the Disjunctive Move, but has since argued against it (1992, 1998) on grounds concerning the confirmation of disjunctive generalizations. The strategy is similar to that of Owens (1989) and Seager (1991).

[7] The Disjunctive Move may arise in almost any potentially reductive context,

e.g. in the context of trying to reduce classical genetics to molecular biology (See Kitcher, 1984, p. 345). The disjunctive move Kitcher considers is not exactly the one that interests us here since the reductionist in his story constructs a disjunction including only all *actual* molecular correlates for genes, not all nomologically possible ones.

[8] It is implicit, for instance, in Putnam (1975), pp. 134–137, and elsewhere.

[9] Cf. Block (1997), pp. 120–121, and Fodor (1997), p. 156.

[10] Cf. Fodor (1997), pp. 156–157.

[11] Pereboom and Kornblith (1991), p. 126.

[12] Cf. Van Fraassen (1980).

[13] Putnam (1981), p. 206. Cf. (1975), p. 137.

[14] Kitcher (1984), p. 347.

[15] Pereboom and Kornblith (1991), p. 127.

[16] An argument along these lines is suggested by Kitcher, *op. cit*., pp. 348–349.

[17] Pereboom and Kornblith (1991), *op. cit*.

[18] Goldman (1989), Gordon (1986).

[19] Translations from the Greek are my own unless otherwise noted.

[20] Kitcher (1984), pp. 347 and 370, is representative here.

[21] Marras (1993), p. 217 makes precisely this sort of claim.

[22] Fodor (1974) was the first to advance this argument. *Pace* Pereboom and Kornblith (1991), he did not advance the Explanatory Response, nor *pace* Antony and Levine (1997), did he advance an argument based on the supposed unprojectibility of disjunctive predicates.

[23] Sklar (1967), Causey (1977), Kim (1998), Ch. 4.

[24] On upgrading biconditionals to identities see Feigl (1958), and Smart (1962). For criticisms of this approach see Brandt (1960), pp. 63–64, Kim (1966), and Lewis (1966, 1972).

## REFERENCES

Antony, L. and Levine, J. (1997): 'Reduction with Autonomy', in J. Tomberlin (ed.), *Philosophical Perspectives 11, Mind, Causation, and World*. Malden, MA: Blackwell, pp. 83–105.

Bealer, G. (1994): 'Mental Properties', *Journal of Philosophy* 91, 185–208.

Block, N. (1980): *Readings in Philosophy of Psychology*, 2 vols, Cambridge, MA: Harvard University Press.

Block, N. (1990): 'Can the Mind Change the World?' in G. Boolos (ed.), *Meaning and Method: Essays in Honor of Hilary Putnam*, Cambridge: Cambridge University Press.

Block, N. (1997): 'Anti-Reductionism Slaps Back', in J. Tomberlin (ed.), *Philosophical Perspectives 11, Mind, Causation, and World*. Malden, MA: Blackwell, pp. 107–132.

Block, N. and Fodor, J. (1972): 'What Psychological States Are Not', *Philosophical Review* 80, 159–181. Reprinted with revisions by the authors in Block (1980), pp. 237–250.

Brandt, R.B. (1960): 'Doubts about the Identity Theory', in S. Hook (ed.), *Dimensions of Mind*, New York: New York University Press, pp. 57–67.

Causey, R.L. (1977): *Unity of Science*, Dordrecht, Holland: D. Reidel Publishing Company.

Davies, M. and Stone, T. (1995): *Folk Psychology*, Oxford: Blackwell.

Feigl, H. (1958): 'The Mental and the Physical', in H. Feigl et al. (eds.), *Minnesota Studies in the Philosophy of Science, Vol. 2*, Minneapolis, MN: University of Minnesota Press, pp. 370–497.

Fodor, J. (1974): 'Special Sciences, or The Disunity of Science as a Working Hypothesis', *Synthese* 28, 97–115. Reprinted in Block (1980), pp. 120–133.

Fodor, J. (1997): 'Special Sciences: Still Autonomous After All These Years', in J. Tomberlin (ed.), *Philosophical Perspectives 11, Mind, Causation, and World*. Malden, MA: Blackwell, pp. 149–164.

Goldman, A.I. (1989): 'Interpretation Psychologized', *Mind and Language* 4, 161–185. Reprinted as Chapter 3 of Davies and Stone (1995).

Gordon, R.M. (1986): 'Folk Psychology as Simulation', *Mind and Language* 1, 158–171. Reprinted as Chapter 2 of Davies and Stone (1995).

Hull, D. (1974): *Philosophy of Biological Science*, Englewood Cliffs, NJ: Prentice-Hall, Inc.

Kim, J. (1966): 'On the Psycho-Physical Identity Theory', *American Philosophical Quarterly* 3, 227–235.

Kim, J. (1972): 'Phenomenal Properties, Psychophysical Laws, and the Identity Theory', *Monist* 56, 177–192. Selections reprinted in Block (1980), pp. 234–236.

Kim, J. (1978): 'Supervenience and Nomological Incommensurables', *American Philosophical Quarterly* 15, 149–156.

Kim, J. (1984): 'Concepts of Supervenience', *Philosophy and Phenomenological Research* 45, 153–176. Reprinted in Kim (1993), pp. 53–78.

Kim, J. (1990): 'Supervenience as a Philosophical Concept', *Metaphilosophy* 21, 1–27. Reprinted in Kim (1993), pp. 131–160.

Kim, J. (1992): 'Multiple Realization and the Metaphysics of Reduction', *Philosophy and Phenomenological Research* 52, 1–26. Reprinted in Kim (1993), pp. 309–335.

Kim, J. (1993): *Supervenience and Mind*, Cambridge: Cambridge University Press.

Kim, J. (1998): *Mind in a Physical World*, Cambridge, MA: MIT Press/Bradford Books.

Kitcher, P. (1984): '1953 and All That: A Tale of Two Sciences', *Philosophical Review* 93, 335–373.

Lepore, E. and Loewer, B. (1989): 'More on Making Mind Matter', *Philosophical Topics* 17, 175–191.

Lewis, D. (1966): 'An Argument for the Identity Theory', *Journal of Philosophy* 63, 17–25.
Lewis, D. (1972): 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50(3), 249–258. Reprinted in Block (1980), pp. 207–215.
Marras, A. (1993): 'Supervenience and Reducibility: An Odd Couple', *Philosophical Quarterly* 43, 215–222.
Mayr, E. (1969): 'Scientific Explanation and Conceptual Framework', *Journal of the History of Biology* 2, 123–128.
Owens, D. (1989): 'Disjunctive Laws?' *Analysis* 49, 197–202.
Pereboom, D. and Kornblith, H. (1991): 'The Metaphysics of Irreducibility', *Philosophical Studies* 63, 125–145.
Putnam, H. (1967): 'Psychological Predicates', in W.H. Capitan and D.D. Merrill (eds.), *Art, Mind, and Religion*, University of Pittsburgh Press, pp. 37–48. Reprinted as 'The Nature of Mental States' in Block (1980), pp. 223–231.
Putnam, H. (1970): 'On Properties', in N. Rescher et al. (eds.), *Essays in Honor of Carl G. Hempel*, Dordrecht, Holland: D. Reidel.
Putnam, H. (1975): 'Philosophy and Our Mental Life', in N. Block (ed.), *Readings in Philosophy of Psychology*, 2 vols, Cambridge, MA: Harvard University Press, pp. 134–143.
Putnam, H. (1981): 'Reductionism and the Nature of Psychology', in J. Haugeland (ed.), *Mind Design: Philosophy, Psychology, Artificial Intelligence*, Montgomery, VT: Bradford Books, pp. 205–219.
Seager, W. (1991): 'Disjunctive Laws and Supervenience', *Analysis* 51, 93–98.
Sklar, L. (1967): 'Types of Intertheoretic Reduction', *British Journal for the Philosophy of Science* 18, 109–124.
Smart, J.J.C. (1962): 'Sensations and Brain Processes', in V.C. Chappell (ed.), *The Philosophy of Mind*, Englewood Cliffs, NJ: Prentice-Hall, Inc., pp. 160–172.
Tomberlin, J. (ed.) (1997): *Philosophical Perspectives 11, Mind, Causation, and World*, Malden, MA: Blackwell.
Van Fraassen, B.C. (1980): *The Scientific Image*, Oxford: Clarendon Press.
Waters, C.K. (1990): 'Why the Anti-reductionist Consensus Won't Survive the Case of Classical Mendelian Genetics', in A. Fine, M. Forbes and L. Wessels (eds.), *PSA 1990*, Ann Arbor, MI: Edwards Brothers, pp. 125–139.

*University of Notre Dame, USA*
*E-mail: jaworski.9@nd.edu*