

## **The Metaphysics of Irreducibility**

Derk Pereboom and Hilary Kornblith

During the 'sixties and 'seventies, Hilary Putnam, Jerry Fodor, and Richard Boyd, among others, developed a type of materialism that eschews reductionist claims.<sup>1</sup> In this view, explanations, natural kinds, and properties in psychology do not reduce to counterparts in more basic sciences, such as neurophysiology or physics. Nevertheless, all token psychological entities--states, processes, and faculties--are wholly constituted of physical entities, ultimately out of entities over which microphysics quantifies. This view quickly became the standard position in philosophy of mind, and reductionism fell out of favor. Recently, however, reductionism has been experiencing a rebirth, and many have suggested that the non-reductive approach was accepted too quickly and too uncritically. In this paper, we attempt to provide a more thorough account of the anti-reductionist position, and, in the process, to defend it against its recent critics.

### I. Irreducibility, multiple realizability, and explanation.

When Putnam first argued for nonreductive materialism, he cited the phenomenon of multiple realizability as its main justification. Since mental states can be realized by indefinitely many neurophysiological states, and perhaps by many non-neurophysiological states, mental states are not reducible to neurophysiological states. Perhaps because the phenomenon of multiple realizability played such a prominent role in Putnam's presentation of this view, many philosophers seem to identify the claim that mental states are not reducible to neurophysiological states with the claim that mental states are multiply realizable. But this is a mistake. We shall argue that multiple realizability is not the only metaphysical feature of irreducibility, and moreover, not the most fundamental feature.

The phenomenon of multiple realizability also played a prominent role in Fodor's general account of anti-reductionism in "Special Sciences." Consider a law in some special science:

$S_1x$  causes  $S_2x$

where  $S_1$  and  $S_2$  are natural kind-predicates in that science. The most appropriate model of reduction requires that every kind that appears in this law be identified with a kind in the reducing science, in virtue of bridge principles. Bridge principles either translate kind-predicates in one science into those of a more basic one, or specify a metaphysical relation, like being identical to or being a necessary and sufficient condition for, between the kinds of one science and those of the reducing science. But in some cases, the sort of bridge principle required for reducibility will not be available.

If kinds in psychology are multiply realizable in an indefinite number of ways at the neurophysiological level, purported bridge principles for relating psychological to neurophysiological kinds will involve open-ended disjunctions. Such purported bridge principles will have be of the form:

$P_1 = N_1 \vee N_2 \vee N_3 \dots$

which says that a certain psychological state,  $P_1$ , is identical to an open-ended disjunction of neurophysiological states,  $N_1 \vee N_2 \vee N_3 \dots$ , or

$P_1 \leftrightarrow N_1 \vee N_2 \vee N_3 \dots$

which says that a certain psychological state is necessary and sufficient for an open-ended disjunction of neurophysiological states.<sup>2</sup> Fodor argues that since open-ended disjunctions of kinds in neurophysiology are not natural neurophysiological kinds, psychological kinds cannot be reduced to neurophysiological kinds.

Why are such disjunctions not natural kinds? Fodor's reason is that they are not natural kinds because they cannot appear in laws. They cannot appear in laws because "laws" involving such disjunctions are not explanatory. Finally, Fodor says that such laws are not explanatory because they do not meet our interests in explanation. Fodor's argument for irreducibility, then, appeals to the fact that purported explanations for psychological phenomena are unsatisfying when couched in terms of open-ended disjunctions. In advancing this claim, Fodor was echoing a

point of Putnam's.

There can be little doubt that Putnam and Fodor are right about this. 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.

For many, however, invoking our interests and the satisfaction of our feelings about explanation seems uncomfortably subjective. The reductionist might say that there exist open-ended disjunctive but nevertheless genuine laws and explanations, even though they might fail to meet certain subjective requirements. 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.

The apparent subjectivity of these anti-reductionist considerations is, however, called into question by an argument for scientific realism. The reductionist wishes to claim that our interests in explanation, which are not met by "explanations" and "laws" couched in terms of open-ended disjunctions, are somehow merely parochial; they reflect parochial interests or limitations of scientific investigators. When it is these interests, however, which give rise to and define successful scientific research programs, the claim that these interests are merely parochial loses its plausibility. The success of a scientific research program in prediction and technological application is evidence of the truth of the theories which are instrumental in gaining that success, and of the legitimacy of the interests which give rise to and define that program. More precisely, it is evidence that the interests which define the standards of explanation which are in part constitutive of that research program are not merely parochial, but instead have a purchase on

objectivity. Our interests in explanation are not objective merely because they are ours, nor is an explanation a good one merely because it satisfies our interests, whatever they might be. Rather, our interests in explanation make a legitimate claim to objectivity when they are instrumental in giving rise to a successful research program.

One might object that although science should not allow laws containing open-ended disjunctions, metaphysics need not be restricted in this way. Hence laws containing open-ended disjunctions might well provide a metaphysical explanations, and hence reductions of psychological laws. We believe, however, that such an attempt to separate science from metaphysics is misguided. The picture invoked by this objection is of an a priori metaphysics, unguided by science. The more appealing view is that since metaphysics and science both aim to characterize and explain the structure of reality, they should not be viewed as separate enterprises. Rather, each should appeal to the other when appropriate. In this particular case, since the notion of a law is a paradigmatically scientific notion, metaphysics should yield to science for its criteria for what counts as a law.

The basis, then, for the claim that psychology is not reducible to neurophysiology is not simply that mental states are multiply realizable at the neurophysiological level, but rather that this multiple realizability shows that attempts at reduction would require laws and explanations of a very peculiar kind; so peculiar, indeed, that they would be unsatisfying as laws and explanations. This dissatisfaction does not rest on merely subjective interests. The legitimacy of our interests depends on the case that can be made for the success of the research program which they partially give rise to and define. We believe that such a case can be made for psychology, and it is there that the argument against reductionism is ultimately founded.

## II. Irreducibility and constitutional explanations.

More, however, remains to be said about the metaphysical state of affairs that obtains when one causal explanation fails to reduce to another. In order to see this, we must first

distinguish between two types of explanation.<sup>3</sup> In addition to the usual causal explanations, we wish to speak of constitutional explanations. When we provide a constitutional explanation for something, we attempt to say what that thing is made of, to specify its constitution. In investigating the issue of reduction, we must look at the different roles constitutional explanations can play.

The notion of reduction at issue in the debate over nonreductive materialism is that of one type-level causal explanation reducing to another. That is, whether reductionism in a certain area is true depends on whether an explanation which quantifies over types or kinds at one scientific level reduces to an explanation which quantifies over types or kinds at another. An exploration of the relationships between such type-level causal explanations and constitutional accounts of the kinds and processes referred to in these explanations will serve to elucidate the metaphysics of irreducibility.

Consider the following type-level causal explanation:

Raising the temperature of the gas in a hot air balloon causes it to rise.

Here, a constitutional account of temperature of a gas as mean molecular kinetic energy allows us to deepen this causal explanation, since greater mean molecular kinetic energy is intimately tied, at the type-level, to lower density of the gas in the balloon, which in turn explains the propensity of the balloon to rise.<sup>4</sup> The constitutional account of gas temperature thus invokes properties that illuminate the above type-level causal explanation. In this case, a constitutional account provides us with a reduction of a type-level causal explanation because the constitutional account illuminates this explanation. Of course, such illumination is not provided by the constitutional account all by itself, but in conjunction with the system of laws and explanations that govern the specified constitution in its particular scientific domain.

When, however, type-level causal explanations are not reducible to explanations in a more basic science, these causal explanations are not illuminated by constitutional accounts in that more basic science. Consider what Philip Kitcher tells us about the relationship between genetics and

molecular biology.<sup>5</sup> In classical genetics, the transmission of genes is accounted for by meiosis, a process in which paired entities "are separated by a force so that one member of each pair is assigned to a descendent entity". Kitcher points out that these processes are not a natural kind from the molecular point of view. The power to separate the paired entities is multiply realizable; all that matters is that the bonds between the originally paired entities be somehow broken. New bonds are sometimes formed between the constituent molecules of these entities, but many accessory molecules may also be involved. Separation may even result from electromagnetic forces, nuclear forces, or gravity. Moreover, separation may occur due to different varieties of these types of forces.<sup>6</sup> In this example, at the token-level, a constitutional explanation tells us how a token genetic process is realized in molecular material. But since a genetic process-type is multiply realizable at the molecular level, constitutional accounts of various instantiations of this genetic process-type are heterogeneous. Consequently, there is no molecular constitutional explanation which illuminates, rather than obscures, the type-level genetic process. In general, because molecular constitutional accounts of genetic process-types fail to illuminate type-level genetic causal explanations, type-level genetic causal explanations are not reducible to molecular explanations.

This is not to say, however, that token-level constitutional accounts are of no interest when there is no reduction in the offing. In all but the special case of microphysical processes<sup>7</sup>, constitutional accounts of token processes will provide illumination. After all, every process is realized in microphysical stuff, and so there is some account to be given of just how it is that the process is so realized. It is one thing to explain how a token process is realized in physical material, but quite another to show that a certain type of process reduces to a physical process. The first is possible for every existing process; the second only for those where there is a constitutional explanation at the type-level.

What is the relationship between the causal powers appealed to in different levels of a constitutional explanation? The answer to this question depends upon whether the constitutional

account under discussion provides us with a reduction. In the cases where a constitutional account does provide us with a reduction, the account allows us to identify the types of causal powers of the objects and properties at the two levels of explanation. In cases where a constitutional account does not provide a reduction, the metaphysical state of affairs is different. Let us return to the example of the relationship between genetics and molecular biology. A type of gene has certain causal powers, active in any type of genetic process, for example, the power to bring about certain traits in the descendants it produces. These causal powers, we shall argue, are not to be identified with the causal powers of the molecules which wholly constitute the gene. Rather, the relationship is, again, one of token-constitution. The following condition expresses our notion of the token-constitution of causal powers:

The causal powers of a token of kind F are constituted of the causal powers of a token of kind G just in case the token of kind F has the causal powers it does in virtue of its being constituted of a token of kind G.

A token gene has the causal powers it does in virtue of the causal powers of the molecules which constitute it; its causal powers do not arise from nowhere. Hence, by our condition, the causal powers of a token gene are constituted from the causal powers of its constituent molecules.

We believe that where there is irreducibility of explanation, there is only token constitution, and no identity of causal powers. In fact, when one type of explanation does not reduce to an explanation at a more basic level, the causal powers at the higher level of explanation are neither type- nor token-identical to causal powers at the more basic level. Let us first consider the thesis for type-identity. Suppose, for purposes of argument, that psychological explanation cannot be reduced to physical explanation. To the psychological state-type, desire for ice-cream, we attribute the causal power to cause ice-cream securing behavior. This type of causal power is not identical to any physical causal power, because it is physically multiply realizable.

One might challenge the existence of types of psychological causal power that are not identical to physical causal powers by denying that types of entities in irreducible sciences other

than microphysics have any causal powers. But this is an unpromising strategy. We naturally attribute causal powers not only to types of psychological states, but to biological types, like genes and bodily organs, and the irreducibility of such types has been successfully argued.

But now, let us examine the token case. Is the token causal power of your present desire for ice-cream, D, identical to the token physical causal power, P, which constitutes it?<sup>8</sup> No. Suppose that P is a token causal power of the molecules of token brain cell B, and that you ingested the molecules of B while eating your favorite baby cereal on the morning of your first birthday. Suppose also that your mother had the choice of two type-identical boxes of this baby cereal on the previous day, and that she chose the one on the left. If she had instead chosen the box on the right, you would not have had token brain cell B (although you would have had a brain cell B' of the same type), and because you would lack this token brain cell, you would also lack its attendant token molecular causal power P (although B' would have had a molecular causal power of the same type).

Nevertheless, you clearly would have had the very same token desire for ice-cream with its attendant token causal power D. Token psychological states, and their causal powers, are individuated by the psychological causal networks in which they play a role. They are not individuated by their physical constitution.<sup>9</sup> Thus, a token psychological state and its causal power can remain the same even when its token molecular constitution, and thus its token molecular causal power, is altered.<sup>10</sup> This result can easily be generalized; hence, when one type of explanation does not reduce to another, there is neither type- nor token-identity between causal powers.

Here, in a diversity of causal powers of natural kinds or in explanations, we encounter the most fundamental metaphysical feature of irreducibility. Multiple realizability is indeed a significant feature of irreducibility because it shows that attempts at reduction would require "laws" involving open-ended disjunctions of heterogeneous kinds. But the deeper metaphysical state of affairs underlying such a situation is the existence a diversity of causal powers at the two

levels of explanation. Multiple realizability is a significant metaphysical feature of irreducibility, but only because it is very powerful evidence of a more fundamental feature, that the causal powers invoked in a lower-level explanation are not identical to those in the explanation which is the target of the attempted reduction.

### III. Davidson's anomalous monism.

There is an interesting relationship between the view we advocate and Davidson's anomalous monism which, we believe, casts light on both these views.<sup>11</sup> Like us, Davidson rejects any reduction of the psychological to the physical. Unlike us, he also rejects the existence of psychological laws. Finally, and also unlike us, he embraces a token-identity thesis. We believe that the differences between Davidson's position and ours turn on our different accounts of causation: Davidson is committed to a Humean account of causation, while we are committed to a non-Humean account.<sup>12</sup>

As Davidson has pointed out, there can be no exceptionless laws in psychology. The reason for this is quite simple. Psychological events do not constitute the whole of reality. For any psychological process one might name, there is always the possibility that some event from outside the sphere of the psychological might interfere with the normal working of that process. Thus, if it were a law of psychology that creatures in mental state  $P_1$  are caused to go into mental state  $P_2$  when stimulated in way  $S$ , it is certainly compatible with this law that a creature in state  $P_1$  who is also stimulated in way  $S$  might suffer brain damage in a car accident before this lawful process has the chance to produce mental state  $P_2$ . Since extraneous phenomena like car accidents are not governed by psychological laws, the psychological laws there are, if any, could not possibly be exceptionless.<sup>13</sup>

We believe that this provides no reason to reject genuine, even if not exceptionless psychological laws, nor the view that they invoke causal powers whose operation is subject to interference. But notice that this is not a position which a Humean about causation can hold.

Since Humeans hold that a statement of the form 'A causes B' is true only if A-like events are always followed by B-like events, the very idea of a statement which is both a causal law and admits of exceptions is self-contradictory. For a Humean to acknowledge the point above about the limited scope of psychology-- that there may be non-psychological events which interfere with psychological processes--forces him to say, as Davidson does, that there are no psychological laws.

Hume's own discussion of causation is devoted to rejecting the robust causal powers we wish to invoke--talk of which is not merely a picturesque way of speaking about exceptionless laws linking events. Once one follows Hume's lead, one is forced to deny the existence of causal laws, not only in psychology, but in any science other than, at most, microphysics. Genetic processes, for example, are subject to interference at the microphysical level, and hence genetic laws will be not be exceptionless. Should we conclude that there is no genetic causation? Our best current science would suggest otherwise. Causal claims in genetics and the rest of the special sciences are spared, however, if we assume the existence of real causal powers, talk of which is not merely reducible to talk of exceptionless laws.

The second disagreement we have with Davidson, on the alleged token-identity of mental and physical events, also hinges on the dispute about the Humean view of causation. We hold that token mental states are physically constituted, but not identical to the token physical states which constitute them. How is it that we distinguish between token mental states and their physical constitutions? We do so, ultimately, on the basis of their causal powers. But, once again, this kind of appeal to causal powers is not open to a Humean about causation. Because the only causal laws there are for a Humean are exceptionless laws, and because there are no causes which are not backed by exceptionless causal laws, there can be no causal powers at the psychological level of description. The only room there might be for genuine causal powers, on the Humean account, is at the microphysical level, for it is only at this level that we may eliminate the possibility of interfering factors. Such a view would leave us with no way to distinguish

between a token mental event and the physical state which constitutes it. Thus, we would thereby be committed, as Davidson is, to a token-identity thesis about the relationship between the mental and the physical.

Davidson's position may thus be derived from ours by replacing our commitment to a non-Humean account of causation with a commitment to a Humean account. Alternatively, some of the characteristic and troubling theses of Davidson's anomalous monism may be avoided by rejecting the Humean view of causation, and by accepting our version of non-reductive materialism.<sup>14</sup>

#### IV. Species-specific reduction and the charge of dualism.

We are now in a position to consider several of the objections to non-reductive materialism.

Kim argues that anti-reductionists assume that even though psychology does not reduce to neurophysiology, there is nevertheless strong connectibility between psychological and neurophysiological states.<sup>15</sup> That is,

(S) For each system and any psychological state that it can instantiate, there is a physical condition of that system which is necessary and sufficient for the system to instantiate that psychological state.

So anti-reductionists assume,

$$S_i \rightarrow (P_i \leftrightarrow M)$$

"which says that for organisms belonging to species  $S_i$  (or systems of physical structure  $S_i$ ) a certain physical state,  $P_i$ , exists which is both necessary and sufficient for the given mental property  $M$ ."

Kim aims to show that the phenomenon of multiple realizability is easily accommodated within a reductionist framework. Thus he states,

What is important then is that these laws are relative to physical-biological structure-types,

although for simplicity I will continue to put the matter in terms of species. The substantive theoretical assumption here is the belief that for each psychological state there are physical-biological structure types, at a certain level of description or specification, that generate laws of this form. I think an assumption of this kind is made by most philosophers who speak of multiple realizations of psychological states, and it is a plausible assumption for a physicalist to make. [38]<sup>16</sup>

In support of this last point, Kim quotes Ned Block: "Most functionalists are willing to allow...that for each type of pain-feeling organism, there is (perhaps) a single type of physical state that realizes pain in that type of organism." [38]

This point of Block's fails to support the claim that Kim is making, and for two different kinds of reason. First, even if there is a single type of physical state that normally realizes pain in each type of organism, or in each structure-type, this does not show that pain, as a type of mental state, is reducible to physical states. Reduction, in the present debate, must be understood as reduction of types, since the primary object of reductive strategies are explanations and theories, and explanations and theories quantify over types. Furthermore, psychological theories and explanations quantify over types of mental states, like pain, that are instantiated by organisms of many different species and structural types. The suggestion that there are species-specific reductions of pain results in the claim that pains in different species have nothing in common.<sup>17</sup> But this is just a form of eliminationism. If we generalized this view to other mental states, then we would be forced to reject the legitimacy of psychological theories that quantify over mental states which are instantiated in more than one species. But this is only part of the difficulty here, for Block's comment about pain does not apply to mental states generally. We agree with Block that there are mental states for which, in the normal case, one will likely find a physiological structure for each species which subserves that state. Given the common genetic heritage of members of the same species, it would be astounding if there were not some commonalities of structure across individuals. The closer one gets to the receptors, the more likely one is to find

such commonalities. For example, when you and I detect that there is one object occluding another in the center of our visual fields, the mechanism by which we detect the edges of these objects is very much the same. But it is not at all plausible to move from this point about structures close to the sensory receptors to a claim about mental states generally.

For familiar reasons, there is no plausibility to the claim that when you and I believe that the GNP of the United States is greater than that of Haiti, there is a single physical structure which underlies that belief in each of us. The further we move away from the sensory receptors, the more unlikely we are to find common physical structures underlying our mental states. The prospects for species-specific reduction here are non-existent. Kim's suggestion then that those who have argued against reduction on the grounds of multiple realizability must grant the existence of species-specific reductions is mistaken. There are good reasons to believe that we will not find single structures subserving particular beliefs, not only across species, but across individuals, and indeed, within individuals across times.

Moreover, the fact that in order to get any species-specific reductions we must limit ourselves to normal cases further undermines the attempt at reduction. When the physical equipment which normal members of species have is damaged and replaced with a prosthesis, the very same mental states are subserved by different physical states. A human being with an artificial eye could detect the edges of objects as well as one with the kind of visual equipment common to the rest of us; we would not want to deny that edge detection occurs in such a person because the physiological laws which cover edge detection in other humans fail to accommodate this particular case. Edge detection is simply not a physiological phenomenon. It is, in the typical case, subserved by a characteristic physiological structure, but it is a higher-level phenomenon governed by higher-level laws. Any attempt to reduce edge detection to the physical mechanisms which standardly (in humans) serve to realize it misses the generalizations which make edge detection the kind of phenomenon it is, namely, a bit of visual information processing.

Admittedly, Kim prefers physical structure-specific, rather than species-specific reduction.

But here again, he must answer the question "What do all of the structure-specific pains have in common?" He would appear to lack resources for a positive answer, and he therefore seems to be committed to eliminating pain as a single type of mental state. Furthermore, since Mary and Jane may have different physical structures realizing their mental states, and since Mary may have different physical structures realizing her mental states at different times, structure-specific reduction has the consequence that there may be nothing that Mary's pain and Jane's pain have in common, and even that there may be nothing that two of Mary's pains have in common. Thus, even if Mary's and Jane's pains, or Mary's pains at different times, are caused by the same types of perceptions, have the same relations to other mental states, cause the same type of behavior, and have the same qualitative characteristics, they would not necessarily be governed by the same laws. Such a proposal certainly eliminates psychological states and psychological explanations as our best psychological theories construe them.

Kim concludes that "the multiple realizability of the mental has no anti-reductionist implications of great significance; on the contrary, it entails, or at least is consistent with, the local reducibility of psychology..." [39] We hope that it is now clear that this view is mistaken. Species-specific correlations are the exception rather than the rule with mental states; they occur only in cases of mental states realized in equipment very close to the sensory receptors. Furthermore, neither species- nor structure-specific reduction is a genuine reduction of psychology, for the attempt to reduce mental talk to talk of species- or structure-specific states eliminates those states to which our best current psychological theories are committed.<sup>18</sup>

A related argument against nonreductive materialism, based on an example from physics, has been advanced by Berent Enç, Patricia Churchland and Paul Churchland.<sup>19</sup> Temperature is clearly a natural kind in physics, while, it is claimed, it is both multiply realizable by completely different physical kinds and reducible to these various realizations. Thus, by analogy, pain may be a natural kind, and although it is multiply realizable in completely diverse neurophysiological states, it may still be reducible to these realizations. Enç and the Churchlands maintain that

according to contemporary physics, temperature in a gas is reducible to the mean molecular kinetic energy of its constituents, while temperature in other media, like a vacuum or a plasma, cannot be realized by mean molecular kinetic energy. Temperature in a vacuum, for instance, is identical to the blackbody distribution of the vacuum's transient radiation. Hence, analogously, the multiple realizability of mental states in thoroughly diverse neurophysiological media is consistent with their reducibility to neurophysiological states.

First, mental states are disanalogous to temperature in a way which undercuts the force of this example. Whereas a particular kind of temperature, such as temperature in gas, can be realized in only one way, belief in a human, for example, can be realized in many ways. Temperature in a gas can only be realized as the kinetic energy of molecules, and is thus not itself multiply realizable, whereas human belief can be realized in many neurophysiological media. Thus, although there is a particular set of laws for gas temperature, couched in terms of mean molecular kinetic energy, human psychological laws cannot analogously be cashed out as neurophysiological laws. Consequently, even if temperature were reducible to its realizations in several domains, mental states in general would not be.

But furthermore, if the various realizations of temperature genuinely had nothing in common, there would be reason to eliminate talk of temperature from our science entirely, rather than maintain the use of the term while simultaneously insisting that it refers to a heterogeneous class. There must be some single property in virtue of which these different realizations all count as temperature, for otherwise they could not qualify as realizations of a single natural kind. Were there nothing these various physical states had in common, the supposed reductions would provide an elimination of temperature. But in fact there is no elimination here. Thermodynamics supplies a characteristic that these different manifestations of temperature do have in common, and it is provided by the following definition:

Temperature is a quantity which takes the same value in two systems that are brought into thermal contact and are allowed to come to equilibrium.<sup>20</sup>

(Ideally, two systems are brought into thermal contact with each other when they are separated by a diathermic wall, a wall which allows exchange of electromagnetic and mechanical forces, but no material exchange.) The systems to which this definition of temperature applies may be homogeneous or heterogeneous pairs of the different media to which Enç and the Churchlands appeal; gasses, liquids, solids, plasmas, and vacuums. Systems of all of these types can be brought into thermal contact and allowed to come to equilibrium. Thus, for example, the temperature of some particular gas and the temperature of some particular vacuum can both be characterized as a quantity which takes the same value when systems made up of the gas and the vacuum are brought into thermal contact and are allowed to come to equilibrium. Hence there is a single characteristic, shared by the various realizations of temperature, to which temperature is reducible. Consequently, Enç and the Churchlands have not produced an example in which a natural kind is realizable by states that have nothing in common, and is yet reducible to these states. Thus they have not provided leverage against the view that the multiple realizability of mental states is powerful evidence for nonreductive materialism.

Finally, Kim claims that nonreductive materialism embraces a dualism about psychological and physical attributes. This claim is misleading. Under no classification is the anti-reductionist's position a kind of dualism. Rather, it is a pluralism at one ontological level, and a monism at the most fundamental level. It is a pluralism in that it holds that there are kinds of entities at many levels of scientific description, and these different levels are not reducible to one another. It is monistic in that it maintains that everything that exists is constituted by microphysical particles.

According to nonreductive materialism, the difference between psychological and physical attributes is no deeper than the difference between biological and physical attributes, or the difference between kind predicates in classical genetics and molecular biology.<sup>21</sup> We would not want to say that because classical genetics does not reduce to molecular biology, there are two fundamental kinds of stuff present here. Similarly, the irreducibility of psychology to neurophysiology does not entail any variety of dualism.

## V. The alleged success of some neurophysiological reductions.

Patricia Churchland and Paul Churchland suggest that nonreductive materialism is mistaken because many psychological phenomena seem to go begging for neurophysiological explanations.<sup>22</sup> The psychological effect of drugs and brain lesions, sleep, and fainting are good examples. Shouldn't we accept some form of reductionism to accommodate these examples?

No. The right anti-reductionist position is that in each science, there is a large body of explanations that do not reduce to explanations in a more basic science. This view is consistent with three points about the relation between psychology and more basic sciences.

First, nonreductive materialism is consistent with the view that some phenomena in psychology may be best causally explained in terms of kinds and properties in some more basic science. Consider, by analogy, the relation between classical genetics and molecular biology. Even though one is not reducible to the other, certain kinds of mutations in genes may be best explained in virtue of changes in molecular structure. In psychology, when the feeling of pain is caused by a pinprick, a macrophysical event, the pinprick, explains a biological event, tissue damage, and this biological event explains a psychological event, the feeling of pain. The loss of psychological functioning by someone who has undergone a lobotomy is best explained in terms of neurophysiological damage. Hallucinating upon ingesting LSD is also best explained by a mechanism more basic than the psychological.

In general, these instances of cross-science causation occur because entities referred to in a higher-level science are constituted from entities at various more basic levels. In cases where psychological state types tend to be realized in common physiological structures, changes in the entities at the more basic levels may result in changes at the higher levels. Such changes will be lawlike if types of changes at a more basic level result in types of changes at a higher level. The lobotomy example may be a case of such a state of affairs. This does not mean, however, that the kinds in the higher-level science are to be reduced to the physical structures which typically realize them, for those higher-level kinds may be multiply realizable at the lower level. Consequently,

there may well be cross-science laws that do not reduce to laws in any more basic science.

Schematically, there might be laws of the following sort:

$N_1x$  causes  $P_1x$  ( $N_1$  is a kind-predicate in neurophysiology, and  $P_1$  is a kind-predicate in psychology)

where this law does not reduce to a purely neurophysiological law because kind  $P_1$  is multiply realizable at the neurophysiological level. The nonreductive materialist can safely admit such cross-science laws, because their widespread incidence fails to undermine the view that large bodies of explanations in the special sciences do not reduce to more basic explanations.

Second, anti-reductionism is compatible with the reducibility of some kinds in some special science to kinds in a more basic science. Tiredness might be nothing more than a single type of biological phenomenon (though we doubt it), and perhaps psychological explanations involving tiredness will be illuminated when recast as involving this biological phenomenon. In order for such reductions to be successful, they would clearly have to be more than species-specific. Yet even the reducibility of some psychological kinds to those in a more basic science does not undermine the irreducibility of other, and indeed most psychological kinds, such as beliefs and desires.

Third, anti-reductionism is consistent with the admission that some psychological laws are reducible to neurophysiology. For example

Tiredness typically causes sleep.

is plausibly a psychological law. But perhaps both tiredness and sleep are each types of neurophysiological phenomena, and can be recast as neurophysiological kinds. Furthermore, and more fundamentally, it may be that when 'tiredness' and 'sleep' are replaced with descriptions of these neurophysiological phenomena, the explanation is illuminated, rather than merely obscured. We would be surprised if the Churchlands were correct about this and the neurophysiological understanding to be gained here were anything more than species-specific. But again, even if the Churchlands were right, it would not follow that all or most psychological laws are reducible to

neurophysiology. Indeed, type-level explanations for actions by beliefs and desires are typically not illuminated by substitutions of neurophysiological constitutional explanations for psychological terms.

#### VI. Explanatory exclusion.

According to Kim, anti-reductionism falls to what he calls the problem of explanatory exclusion.<sup>23</sup> Consider a particular bit of human behavior. According to nonreductive materialism, he maintains, there will be two causal explanations for this event. One is physical in nature, another is psychological. There is a physical explanation of the event in virtue of the causal closure of the physical domain: any physical event that has a cause has a physical cause. At the same time, most of us grant that this event has an explanation in terms of beliefs and desires as well.

But how are we to understand the relationship between these different causes? Kim considers two alternatives. One is that they are separately insufficient but jointly sufficient to cause behavior. Each of the two explanations would then yield partial causes of the effect. Kim thinks that not only does this seem absurd, but it also violates the principle of the causal closure of the physical domain. We agree. Such an explanation is absurd, but neither is it a part of non-reductive materialism.

A second possibility is that the effect is overdetermined. If the beliefs and desires had not occurred, the physical causes would still have been sufficient to cause the effect, and vice versa. One reason Kim thinks that this is mistaken is that it is absurd to think that there are two independent causal chains leading to the same effect. From what we know about the physiology of limb movement, we must believe that if the pain sensation causes my hand to withdraw, the causal chain from the pain to the limb motion must somehow make use of the causal chain from some appropriate central neurophysiological event to the muscle contraction. We agree.

Kim points out that the problem would be solved if we said that the mental and the physical cause are one and the same. Indeed, he is right about this. Identifying mental with

physical causal powers is one way of vindicating the causal efficacy of the mental.

Yet the anti-reductionist should not give up this easily. Both genetics and microphysics are paradigmatic physical sciences, since the entities and causal powers of each are clearly physical--that is, constituted of physical stuff. But when we explain the child's having blue eyes by means of genetics, we know that at the same time there is a microphysical explanation for the microphysical states that constitute the child's having blue eyes. Consequently, here we also run into the explanatory exclusion problem; for any genetic event, Kim would have to say that there is both a genetic and a microphysical explanation. The existence of the explanatory exclusion problem for sciences whose entities are manifestly constituted of physical stuff should mitigate Kim's worries about explanatory exclusion in the case of psychology and neurophysiology. There is no special problem about explanatory exclusion for psychology. If there is a problem here at all, it is a problem about the relations among the special sciences generally, not one about the relation between the mental and the physical.

We do not mean to argue, however, that we need not be concerned with this problem because it is ubiquitous. Quite the opposite is true. The solution to Kim's problem, however, is easily provided within the account of irreducibility given above. In rejecting a reduction of mental states to the physical states which realize them, we need not choose between saying that the mental causal powers are insufficiently efficacious to produce behavior on their own (they are only partial causes of behavior) and saying that they are wholly independent of the physical states (and thus behavior is overdetermined). Rather, as we have already indicated, mental causal powers are wholly constituted of physical causal powers; they are neither identical to (nor are they necessary and sufficient for) them, nor wholly independent of them. The psychological explanation of an event does not compete with its physical counterpart because the mental causal powers referred to in the psychological explanation are wholly made up of the physical causal powers referred to in the physical explanation. Hence, the claim that a bit of behavior was caused by certain mental states is not an explanation which competes with the physical account which

underlies it, any more than the claim that I secured ice-cream with cash competes with the claim that I secured ice-cream with bits of paper and metal. Kim fails to take account of this possibility because he does not recognize that non-reductive materialism is committed both to a constitutional account of mental objects and mental properties, and thereby to a constitutional account of mental causal powers.

## VII. Conclusion.

We do not believe that the reasons which have been offered of late for rejecting non-reductive materialism should be accepted. The reasons for which reductive accounts were largely abandoned remain good ones, and non-reductive materialism remains the most satisfying and sensible account of the relationship among the special sciences.<sup>24</sup>

Department of Philosophy

University of Vermont

## NOTES

1. Richard Boyd, "Materialism Without Reductionism: What Physicalism Does Not Entail," in Ned Block ed., Readings in the Philosophy of Psychology. Volume 1. (Cambridge, Mass.: Harvard University Press, 1980); "Scientific Realism and Naturalistic Epistemology," in Proceedings of the Philosophy of Science Association 1980, v. 2; "The Current Status of Scientific Realism," in Jarrett Leplin ed., Scientific Realism, (Berkeley and Los Angeles: University of California Press, 1984); Jerry Fodor, "Special Sciences," in Readings in the Philosophy of Psychology, ed. by Ned Block, (Cambridge, Mass.: Harvard University Press, 1980); Alan Garfinkel, Forms of Explanation: Rethinking the Questions in Social Theory, (New Haven and London: Yale University Press, 1981); Hilary Putnam, "The Nature of Mental States," in Philosophical Papers, volume 2, (Cambridge: Cambridge University Press, 1975), "Philosophy and Our Mental Life" in Philosophical Papers, volume 2.
2. For economy in exposition, we shall focus on reduction in terms of identity, and ignore reduction in terms of necessary and sufficient conditions in the first three sections of this paper.
3. Here we follow Robert Cummins, The Nature of Psychological Explanation, (Cambridge Mass.: Bradford Books/MIT Press, 1983). We have changed Cummins terminology slightly (he speaks of analytic explanations where we speak of constitutional explanations), but there is no disagreement in substance. We are indebted to Cummins throughout this section of the paper.
4. This identification is, in fact, an oversimplification, but in ways which do not effect this particular point. For a more precise account of the relationship between temperature and mean molecular kinetic energy, see section IV below.
5. Philip Kitcher, "1953 and All That: A Tale Of Two Sciences," The Philosophical Review XCIII, No. 3 (July, 1984).
6. Kitcher, pp. 349-50.
7. This is a special case, because fundamental microphysical processes are not composed of anything else at all, and so there are no constitutional accounts to be had.

8. Some may reject talk of token causal powers, and prefer instead talk of the (token) instantiations of causal powers. Our argument is no less effective if couched in these terms.
9. Note how odd it would be to individuate token psychological states by their physical constitution. We often want to know whether someone would continue to have the very same belief given certain changes in his environment or body chemistry. Surely it would be pointless for a psychological theory to count every physical change, however trivial, as a change in the token belief. This contrasts with cases in which the physical change results in the belief being rejected or extinguished and later adopted again, perhaps for different reasons. Here there would be reason to speak of a different token belief, but, again, not for reasons having to do with physical constitution.
10. Here we follow Boyd and Cummins, rather than Fodor. Fodor does not acknowledge that the points he makes against identification of mental state types with neurophysiological types can also be made to show that even mental state tokens should not be identified with neurophysiological tokens, for mental state tokens are also multiply realizable at the neural level.
11. "Mental Events," in Essays on Actions and Events, (Oxford: Clarendon Press, 1980), 207-225.
12. See "Causal Relations," in Davidson, op. cit., 160.
13. Roughly this point is made by Fodor in "Special Sciences," p. 129. See also Louise Antony, "Anomalous Monism and the Problem of Explanatory Force," Philosophical Review XCVIII, No. 2 (April, 1989), p. 176.
14. The importance of a non-Humean account of causation for an adequate defense and elaboration of non-reductive materialism has been pointed out by Boyd, Materialism Without Reductionism: Non-Humean Causation and the Evidence for Physicalism, manuscript.
15. Jaegwon Kim, "The Myth of Nonreductive Materialism," Proceedings and Addresses of the American Philosophical Association, vol. 63, #3 (November, 1989), pp. 31-47. Page numbers in the text are in square brackets.

16. See also Paul Churchland, Matter and Consciousness, Revised Edition, (Cambridge, Mass.: MIT Press, 1988), pp. 38-42; Berent Enc, "In Defense of the Identity Theory," Journal of Philosophy, LXXX, no. 5, (1983).
17. Ned Block, "What is Functionalism?" in Ned Block ed., Readings in the Philosophy of Psychology, volume 1, (Cambridge, Mass.: Harvard University Press, 1980).
18. A further reply to Kim is inspired by the version of anti-reductionism according to which part of what determines whether a subject is in a belief state of a certain type is the nature of physical environment in which that subject is embedded. For instance, Burge argues that holding all of an individual's states, non-intentionally described, fixed, that individual may or may not have a belief about water, depending on whether there is actually water in the environment. (Tyler Burge, "Individualism and the Mental," in Midwest Studies 1978, (Minneapolis: University of Minnesota Press, 1978); "Other Bodies," in Andrew Woodfield ed., Thought and Object, (Oxford: Oxford University Press, 1982)) In this view there is no physical condition of an individual, let alone a species or a structure-type, which is sufficient to instantiate the psychological state. It is left open by Burge's argument, however, whether there is a physical condition of the individual or species-cum-environment, i.e. of the universe, which is sufficient for the instantiation of the psychological state.
19. Berent Enç, "In Defense of the Identity Theory," Journal of Philosophy, 1983, pp. 279-298; Patricia Churchland, Neurophilosophy, (Cambridge, Mass.: Bradford Books/MIT Press, 1986), pp. 356-8; Paul Churchland, op. cit., pp. 41-42.
20. T. J. Quinn, Temperature -- Monographs in Physical Measurement, (London: Academic Press, 1983), p. 3, cf. pp. 3-17; Mark W. Zemansky and Richard H. Dittman, Heat and Thermodynamics, Sixth Edition, (New York: McGraw-Hill, 1981), pp. 3-10; L. D. Landau and E. M. Lifshitz, Statistical Physics, (Oxford: Pergamom Press), pp. 32-4. There are certain precisely circumscribed conditions, which involve cases in which systems do not come to thermal equilibrium in a reasonable amount of time, under which the notion of temperature is ill-defined

(Quinn, pp. 13-15).

21. The illusion that there is a deeper difference is, we believe, in part created by Kim's terminology. Kim speaks of the mental and the physical. By "the physical" he sometimes means the phenomena over which microphysics quantifies. In this sense, we are committed to the existence not only of the mental and the physical, but also the mechanical, the chemical, the biological and so on. But Kim also sometimes uses the term "the physical" to lump together all of the phenomena of the "unproblematically physical" sciences: i.e. particle physics, mechanics, chemistry, biology, etc. We reject the suggestion that the mental is problematic in a way in which these sciences are not.

22. Paul Churchland, op. cit.; Patricia Churchland, Neurophilosophy, (Cambridge, Mass.: Bradford Books/MIT Press, 1986).

23. Kim, pp. 43-47. See also Kim, "Mechanism, Purpose, and Explanatory Exclusion," Philosophical Perspectives, 3 (1989); and "Explanatory Realism, Causal Realism, and Explanatory Exclusion," in Realism and Antirealism (Midwest Studies in Philosophy XII), Peter French et. al., eds. (Minneapolis: University of Minnesota Press, 1988).

24. We wish to thank Lynne Rudder Baker, Stephen Brush, David Christensen, Richard Healey, Jaegwon Kim, George Sher, David Y. Smith, Sydney Shoemaker, and Ken Waters.