

ROBERT M. FRANCESCOTTI

THE NON-REDUCTIONIST'S TROUBLES WITH
SUPERVENIENCE

(Received in revised form 29 May 1996)

Most physicalists today believe that mental properties can be neurologically realized in many different ways. If some brand of functionalism is correct, then an internal event realizes a mental property in virtue of playing the right functional role, and not because of its intrinsic physical features. This allows that events that differ in neural type might realize the same mental property, the only constraint being that each of these event can play the functional role definitive of that mental property.

The belief in multiple realizability is one of the main reasons for the popularity of *non-reductive physicalism*. If mental states are multiply realizable at the neural level, then they are also multiply realizable, and to a far greater degree, at more basic levels of physical structure – e.g., the molecular, atomic, and subatomic levels. Of course, the properties of the physical sciences are also multiply realizable; the property of being a carbon atom, for instance, is multiply realizable with respect to subatomic features. However, the problem with mental properties is that they, unlike chemical properties, are thought to be multiply realizable with respect to the *entire* range of features mentioned by the physical sciences. That is, for any mental property M, and for *any* event of physical type P, it is possible that M is realized by an event of some physical type other than P. If so, and if this possibility is nomological as well as metaphysical, then mental properties are *irreducible* in the sense that

(NR₁) there are no nomologically true biconditionals connecting mental properties with physical properties (i.e., properties of the physical sciences).

Given that property-identity requires biconditionals that are (at least) nomologically true, it would also follow that

(NR₂) there are no true identity statements connecting mental properties with physical properties; in other words, mental properties are not physical properties.

While non-reductive physicalism (and theses NR_1 and NR_2 , in particular) has become the dominant view in the philosophy of mind,¹ it has had its share of criticism. One common objection is that the bridge principles necessary for theoretical reduction need not take the form of biconditionals as NR_1 assumes.² Another worry is whether the proponent of NR_1 and NR_2 can adequately defend mental causation against Kim's (e.g., 1989, 1993) explanation exclusion argument. There is also the debate over whether acceptable bridge principles might be formed via the "disjunctive strategy", that is, by invoking biconditionals of the form ' $M \leftrightarrow (P_1 \vee P_2 \vee \dots \vee P_n)$,' whose disjuncts designate all the various types of physical events that might realize mental property M .

For the present discussion, I shall concede each of these disputes to the non-reductionist because I wish to address the more basic question, "In what sense are the proponents of NR_1 and NR_2 *physicalists*?" Non-reductionists typically express their allegiance to physicalism with two claims: that mental properties are *physically realized*, and that mentality *supervenes* on features of the physical world. I will show that while both claims are essential to physicalism, when taken together they conflict with NR_1 and NR_2 .

1. PHYSICALISM AND PHYSICAL REALIZATION

Physicalists believe that mentality is purely physical. That is, all the facts about mentality are really just facts about the physical world. But if mental properties are not identical with physical properties, in what sense can physicalism be true?

One very natural answer is that while mental properties are not identical with any of those mentioned by the physical sciences, they are nonetheless physical in a broader sense. Although tables and chairs are not mentioned by the physical sciences, they are physical in a more ordinary, but still very real, sense. The same applies to mentality. Suppose our theory of mind refers only to features that are at least as physical, in the ordinary sense, as tables and chairs. Suppose, for example, that it characterizes mentality in terms of dispositions to behave and/or internal computational processes. Aren't we then entitled to view mental properties as genuinely physical

even though they cannot be identified with the items of the physical sciences?

The problem with this line of response is that it is notoriously difficult to specify the criteria for an item being physical in a way that includes common-sense “physical” items, such as tables and chairs, while excluding those items that would be accepted only by dualists. For instance, by defining ‘physical’ as that referred to by physics proper (either current physics or some ideal physics), we certainly exclude Cartesian soul substances from the realm of the physical, but we also leave out the unquestionably physical features of chemistry, molecular biology, and neurophysiology. On the other hand, if we try broadening the conception by appealing to those general features traditionally associated with matter (e.g., being extended, solid, inert, impenetrable, and governed by deterministic laws), then we risk excluding much of what physics itself describes.³

To avoid these definitional problems, one common strategy is to reserve the term ‘physical’ for those features that are unquestionably physical (e.g., those features mentioned by physics, chemistry, molecular biology, and neurophysiology), and then to defend physicalism by showing how mentality *depends* in some intimate way on these paradigmatically physical features.

The question now becomes: *How* does mentality depend on these paradigmatically physical features? Van Gulick (1992) answers, “[i]n every instance in which a property applies to the world of space and time it does so *in virtue of* physical properties that apply to the world of space and time” (p. 164, my underlining). Assuming he means “*solely* in virtue of,” this thesis does seem to capture the physicalist’s position, although we now face the task of characterizing the “solely in virtue of” relation. Van Gulick’s characterization (p. 171) is a common one; mental properties are had solely in virtue of physical properties because while they are not identical with physical properties, they are nonetheless *realized* by them.

Whether this construal of the *in virtue of* relation makes non-reductionism sufficiently physicalistic depends on what it is for a property to be “realized” physically. In the philosophy of mind at least, talk of realization was especially popular amongst the token-identity theorists. Since Putnam’s (1967) appeal to multiple realizability, mental state-*types* were no longer identified with neural

state-types, since it was thought that *tokens* of many different neural state-types might be *tokens* of the very same mental state-type. This “type-token” jargon suggests that realization is a matter of *property instantiation*; to say that a neural state-token has realized a mental state-type is to say that a neural event has instantiated a mental property. So if mental properties are always realized physically, then instances of mental properties (i.e., mental events) will be nothing more than physical events. Isn’t that all physicalism demands?

Unfortunately, it is not. For all that physical realization requires, the types of physical events that realize my mental properties at the present time might instantiate a radically different set of mental properties on some other occasion, or even worse, they might fail to instantiate any mental properties at all. Mere physical realization also allows that an exact physical duplicate of an ordinary paper-clip might have a rich store of thoughts and sensations, the only constraint being that all instances of those thoughts and sensations are instances of whatever physical properties are had by the paper clip. That doesn’t sound very physicalistic at all. Physicalist intuitions tell us not only that mental properties are instantiated by physical events, but also that there are good physicalistic reasons why the duplicate of the ordinary paper clip will not have mentality but my duplicate will. Physical realization is not enough to secure physicalism unless we also assume something about the way in which mental properties *correlate* with physical properties.

Here we find the familiar appeal to *supervenience*. Even though mental properties are not identical with physical properties, it is nonetheless true that changes in mental properties can come about only as a result of changes in physical properties. To express the same sentiment equally roughly: the physical properties that are had determine all the mental properties that are had. These rough formulations leave a variety of distinct supervenience claims to choose from, and which brand we pick depends on how we decide to make the covariance relation more precise.

2. SUPERVENIENCE

A relatively weak brand of supervenience is what Kim calls “global supervenience” (1984(a), p. 168). When applied to mental and physi-

cal properties, global supervenience requires that any two worlds that are physically indiscernible are also mentally indiscernible. Whether two worlds are indiscernible is a matter of how the properties in question are distributed over the individuals of those two worlds. More precisely, two worlds w_1 and w_2 are indiscernible with respect to a class of properties A just in case the following is true: for every property F in A , and for every individual x , x has F in w_1 if and only if x has F in w_2 . So if w_1 and w_2 are physically indiscernible, then for every individual x , x will have all and only those physical properties in w_1 that it has in w_2 . Global supervenience tells us that, in such a case, x will also have the very same mental properties in w_1 that it has in w_2 .

While global supervenience does capture some of our physicalist intuitions, it falls a bit short. If two worlds are physical indiscernible, then they are also mentally indiscernible. However, when this condition is satisfied *trivially* (that is, when two worlds differ physically), global supervenience places no constraints on how mentality is distributed. The result, Kim warns (1987, p. 321), is that if a world w physically differs from the actual world in even the slightest way (e.g., Saturn's rings containing one more ammonia molecule), then w can differ mentally from the actual world as radically as you please. Global supervenience allows, for instance, that mentality is completely absent in w or distributed across its inhabitant in a very different way, despite the only trivial physical difference. So w might contain creatures who are physically indistinguishable from humans, but who have no thoughts at all, and the duplicates of our rocks and paper clips might be chock-full of mentality there.

To avoid these counter-intuitive results, we need to require that physically indistinguishable creatures will have all the same mental features, even when the worlds they inhabit globally differ. This assurance is provided by Kim's "strong supervenience" (1984(a), p. 165). To say that mental properties *strongly* supervene on physical properties is to say that

(S) necessarily, for each individual x and each mental property M , if x has M , then there is a physical property P such that x has P , and necessarily if any individual y has P , it has M .⁴

Unlike global supervenience, S precludes worlds that differ only trivially from the actual world in terms of physical features while

differing greatly in terms of mental features. A world w that physically differs from the actual world only by having one more ammonia molecule cannot lack mentality, since according to S , the individual physical properties that yield mentality in the actual world guarantee that the same mental properties are also had by our physical duplicates in world w . We are also assured that rocks and paper clips in w lack mentality, since these items will lack the physical properties that yield mentality in the actual world.

An advantage that S has over some brands of supervenience is one that it shares with the global variety. S allows the base properties to include *relational* features, and this is important for two reasons. If the presence of a mental property depends not just on individual internal features but also on relations between internal features, then these internal relations must be included among the base properties for supervenience to obtain. For this reason, S is superior to the brand of supervenience Teller (1986) calls “local physicalism,” according to which “all the non-relational properties of an individual a supervene on a ’s physical non-relational properties” (p. 73). Another reason for allowing relational properties in the supervenience base is to honor externalist intuitions regarding mental content. If mental content is partly a function of relations to environmental features (e.g., being suitably causally related to stuff that is H_2O rather than XYZ), then provided that these external features and our causal relations to them count as physical, they too should be included in the supervenience base. In this way, S improves on certain varieties of *mereological* supervenience. According to Kim (1984(b)), mereological supervenience “requires the supervenience of the characteristics of wholes on the properties and relations characterizing their proper parts” (p. 264). Although Kim does not insist on it, if the relations of the proper parts are wholly *internal* (that is, relations that the parts bear only to other parts of the same whole), then externalism cannot be accommodated.⁵

Like many other varieties of supervenience, S does leave open questions about modality, but it is not hard to see what type of modal guarantee the non-reductionist ends up with. Entailments of the form ‘If x has physical property P , then x has mental property M ’ are certainly not logical truths. Nor are they conceptually true, at least if ‘physical’ is restricted to the subject matter of the physical

sciences. Mere conceptual analysis certainly does not reveal, for instance, which patterns of neural activity yield pain.⁶

Could such statements be metaphysically, though non-conceptually, true? Perhaps mental properties are natural kinds in the Kripkean sense. Pain is essentially, though non-conceptually, neural activity X just as gold is essentially, though non-conceptually, stuff comprised of the element Au. Perhaps, but this option is certainly not open to the non-reductionist. If mental properties are Kripkean natural kinds, then both NR₁ and NR₂ are automatically guaranteed to be false. So it seems that the non-reductionist is entitled to nothing stronger than nomological necessity.

The fact that the guarantee is only nomological provides an answer to Jack (1994, p. 432) who argues that the belief in supervenience is not essential to physicalism. Jack notes that “[s]ome materialists seem to claim that materialism is metaphysically necessary, but materialism does not itself require this” (p. 432). If so, then we can be genuine materialists (physicalists) while allowing mentality to be completely unrestrained by physical features in worlds that are only metaphysically possible. Thus, “[m]aterialism does not require that the mental weakly, let alone strongly, metaphysically supervenes upon the physical” (p. 432). Granted, physicalists need not claim that their thesis holds in all metaphysically possible worlds. But, as we have seen, they need not (and should not if they are non-reductionists) interpret the modality of S metaphysically either. At the same time, however, the physicalist will regard the covariance of mental properties with physical properties as more than merely accidental. While the covariance might not be metaphysically necessary, it must at least be grounded in the laws of nature. Otherwise, it is hard to see in what sense mentality is had *in virtue of* physical features.

Moreover, it seems that nothing weaker than a nomic covariance of the sort expressed by S will satisfy physicalist intuitions. If the mental differences between two individuals were not accompanied by *any* physical differences (not even of the relational variety), then mentality would not be had *solely* in virtue of physical features.

This is not to say that S is true. It might be that mental events have certain features (e.g., causal powers) that are underdetermined by physical features. It might also be that these novel causal powers

affect the course of events in the physical world (thereby violating the causal closure of the physical domain). In fact, if Moser (1992) is right, rejecting these novel causal powers is entirely unwarranted, indicating nothing more than a physicalistic prejudice. But the point remains, it is a *physicalistic* prejudice, and one that is essential to the physicalist's position. For to allow that mental events might have novel causal powers (i.e., causal powers that do not supervene on the causal powers of physical events) is to deny that features of mentality are had solely in virtue of features of the physical world.

The problem with S is not that it is inessential to physicalism, but that it is insufficient. Supervenience claims tell us how mental properties covary with physical properties, but as Kim notes (e.g., 1993(a), pp. 167–168), they say nothing about what *grounds* the covariance relation. They are, therefore, perfectly compatible with a variety of dualistic theories. We may endorse supervenience and agree with the epiphenomenalist that instances of mental properties are not instances of physical properties. Supervenience, even the strong variety S, is also perfectly consistent with parallelism; changes in mental properties require changes in physical properties, not because the latter cause the former, but as result of divine intervention.

But suppose we conjoin S with the view that mental properties are always physically realized? Wouldn't we then have a view that differs significantly from dualist theories? And if so, wouldn't we have a position that truly warrants the title "physicalist?" Perhaps, but I wonder what is then left of non-reductionism. Kim has argued (e.g., 1989, sec. IV) that depending on how we augment supervenience theses to explain how the mental is grounded in the physical, we run the risk of psychophysical reduction. There is, however, a more basic worry. Whatever needs to be added to S to make it physicalistic, the fact remains that S is required, and once S itself is granted, there is trouble for the non-reductive physicalist. In the following section, I show that by endorsing S in an effort to preserve physicalism we risk forfeiting NR_1 and NR_2 .

3. SUPERVENIENCE AND NON-REDUCTIONISM

If some brand of functionalism is correct, as most non-reductive physicalists believe, then whether a physical event e realizes (i.e., instantiates) a mental property is not a matter of e 's intrinsic physical features. Which mental property e realizes depends, instead, on how e causal relates to other events – especially, other internal events, sensory input, and behavioral output. So the neural event that now instantiates my current belief that koalas make good pets would have instantiated the desire to visit Milan if it had played whatever functional/causal role characterizes the latter. In fact, the very same type of neural event would not have instantiated any mental property at all if it had played none of the functional roles characteristic of mentality.

The intuition that functional role (rather than intrinsic physical type) is what matters to mentality is precisely what motivates the belief in multiple realizability. If functional role is what determines which mental property a physical event realizes, and if these functional roles can be played by events that differ in intrinsic physical type, then events of different intrinsic physical types can realize the very same mental property. This functionalist intuition may also be expressed in terms of *underdetermination*. The presence of a mental property underdetermines the intrinsic physical features of its realizer. More precisely,

U_1 : for any mental property M , and any intrinsic physical feature P , such that an instance of P realizes M (for some creature at some time), the presence of P is not guaranteed by the presence of M .

The functionalist intuitions that motivate the belief in multiple realizability also imply underdetermination in the opposite direction. If realizing a mental property is a matter of playing the right functional role, then intrinsic physical type will not determine which, if any, mental property is realized. Events of the same intrinsic physical type might realize different mental properties, and perhaps no mental properties at all, depending on how they relate to other events. Thus, mental features are underdetermined by physical features in the following sense:

U_2 : for any mental property M , and any intrinsic physical feature P , such that an instance of P realizes M (for some creature at some time), the presence of M is not guaranteed by the presence of P .⁷

As far as supervenience is concerned, U_2 raises a crucial question. To say that a property A supervenes on a property B is to say that the presence of B guarantees the presence of A. So if U_2 is true, then mental properties do not supervene on the intrinsic physical features of their realizers. On what, then, do mental properties supervene?

Let us widen our focus by considering, not just the physical realizer itself, but also the types of events with which the realizer interacts. Suppose a neural event e yields a mental property M, for some creature at some time, by causally interacting with other neural events $e_2, e_3, \dots e_n$. If we accept the functionalist story, we cannot say that e 's intrinsic neural type guarantees the presence of M. The fact that e is interacting with $e_2, e_3, \dots e_n$ does not, by itself, guarantee M either. What ensures the presence of M is that e interacts with $e_2, e_3, \dots e_n$ *in the appropriate way* – that is, in the manner definitive of M. Thus, if $C(x, x_2, x_3, \dots x_n)$ is the functional/causal relation definitive of M, then what ensures the presence of M is the second-order property: *having events that satisfy the relation $C(x, x_2, x_3, \dots x_n)$* . This functional property is the feature on which M supervenes.

To illustrate, consider the physical property *density*. Density is multiply realizable with respect to atomic make-up in the sense that instances of many different types of atoms can interact to produce the same degree of density. Items whose constituents differ in atomic type might have the same density, the only constraint being that the constituent atoms are arranged into molecules and those molecules bond in such a way as to yield the same ratio of mass to volume. Thus, having high (low) density does not supervene on intrinsic atomic type, but rather on the second-order property, *having an inner structure that yields a high (low) ratio of mass to volume*. Likewise, mental property M does not supervene on intrinsic neurological type, but on the second-order property, *having events that satisfy the relation $C(x, x_2, x_3, \dots x_n)$* . How exactly we decide to characterize this second-order, functional feature (hereafter “F”) depends on which brand of functionalism we choose to endorse. Yet, whatever F involves, the point remains: it is F that determines whether some collection of neural events manages to yield M.⁸

A collection of physical events guarantees the presence of M only by being related F-wise. Since supervenience thesis S requires that

each mental property is guaranteed by a physical property, to remain loyal to S we will have to assume that F itself is a physical property. However, according to functionalism, F is not only sufficient, but also necessary, for M. So there will be a physical property (namely, F) that is both necessary and sufficient for the presence of M. But in that case, the non-reductionist thesis NR_1 will be false. In fact, if a mental property is *identical* with having the associated functional role played (as most functionalists believe), then NR_2 must also be rejected. On the other hand, if we deny that F is a physical property so as to preserve non-reductionism, then we will have to reject supervenience thesis S. After all, F is the property on which M supervenes, and S requires that M supervene on a physical property. However, if we reject S, then it is hard to see in what sense we are physicalists. As shown in section 2, although the truth of S does not guarantee physicalism, it is at least a necessary condition.

So the non-reductive physicalist faces the following dilemma. The question is whether the functional properties essential to mentality are themselves physical properties. If so, then supervenience is preserved, but only at the expense of NR_1 and NR_2 . If not, then NR_1 and NR_2 are preserved, but only at the expense of physicalism.

I imagine the non-reductive physicalist will respond by rejecting the second horn of the dilemma. It might be argued that the functional properties that ensure the presence of mentality are not themselves physical features, but they nonetheless supervene on physical features. In this way, the corresponding mental properties will also supervene on, but remain irreducible to, those underlying physical features.

Although initially plausible, a closer look will show that this line of response is entirely unmotivated.

4. MERE SUPERVENIENCE

The functional property F, which is identical with M, might supervene on some physical feature, or class of physical features, P without being identical with P. If so, then M will supervene on P without being identical with P. This would be one way for supervenience thesis S to be true without M being identical with any physical

feature. The functional property F supervenes on physical features, but it *merely* supervenes.

What might the base property P be in this case? Well, as we saw, what ensures the presence of M at the physical level is not the types of events that interact to yield M, for events of these types might have interacted differently, in which case M would not have been present. What ensures the presence of M is whatever complex physical *relations* hold between individual physical events when they interact to yield M. In other words, the relations between events at the physical level rather than the relata themselves are what guarantee the presence of M. Still, the functional property F is not identical with any of these underlying physical relations, whatever they might be. Again, F *merely* supervenes.

While this is a possible position, it is hard to see what motivates it. Considerations of multiple realizability provide no motivation. They show that for any mental property M, events of different neural types can fill in for 'e' in whatever functional schema ' $(\exists x_2)(\exists x_3) \dots (\exists x_n)C(e, x_2, x_3, \dots x_n)$ ' correctly characterizes M. Since the type of neural event that realizes M can vary from one instance of M to another, M cannot be identified with any one of those neural types. Considerations of multiple realizability also show that the neural events with which e interacts to yield M may vary greatly in type; the only constraint is that these events, whatever their intrinsic neural features might be, are functionally related in the appropriate way. However, what multiple realizability does *not* show is that the type of *interaction* that obtains at the physical level, in virtue of which diverse collections of neural events all manage to yield M, might itself vary from case to case. For all that multiple realizability shows, it may be that all of the various collections of events that interact to yield M do so in virtue of standing in some relation R at the physical level. If so, then the physical property, *having events related R-wise*, would be nomologically necessary as well as sufficient for the presence of M. It would be sufficient, since this relation between physical events guarantees the presence of functional property F, and therefore mental property M. And it would be necessary because unless physical events are related R-wise, F will not be present, and therefore M will not be present. However, since having events related R-wise is a physical property, thesis NR₁ would then be false. It might

even be that F (i.e., M) is *identical* with the physical property, having events related R-wise, in which case NR₂ would also be false. Thus, considerations of multiple realizability do not show that F merely supervenes on interactions obtaining at the physical level.

To make this point less abstract, consider the anatomical property of being a tendon. This property does not supervene on the intrinsic physical features of tissues, since a tissue will not be a tendon unless it plays the right functional role – namely, serving to connect muscle fibers to bones. Moreover, the intrinsic features of tissues matter only to the extent that they enable tissues to play this functional role. So the functional property, being a tissue connecting muscle fibers to bones, is the feature on which being a tendon supervenes. Of course, we do not say that being a tendon merely supervenes on this functional property; we say that it *is* this property. Nor do we think that the functional property, being a tissue connecting muscle to bone, merely supervenes on features obtaining at the physical level; we think that it itself is a genuinely physical feature. Considerations of multiple realizability provide no reason to think otherwise. Tissues of many different intrinsic physical types can be tendons by playing the right functional role; but the most this shows is that being a tendon cannot be identified with any one of those intrinsic physical features. Thus, multiple realizability allows that the property, being a tissue connecting muscle to bone, is identical with some genuinely physical feature (albeit one of a functional variety). Likewise, considerations of multiple realizability provide no reason to think that the functional roles characteristic of mentality merely supervene on physical features.

Another possible reason for thinking that F merely supervenes stems from explanatory considerations. Many have argued that the physical sciences lack the conceptual resources to adequately explain mental phenomena (e.g., Putnam (1975), Fodor (1974), and Van Gulick (1992)). Every instance of a mental property may be nothing more than a physical event, but it does not follow from this that the physical sciences can adequately *explain* the nature of mentality. One worry is that the accounts offered by physics and chemistry, for example, fail as explanations of mental phenomena for the same reason that they fail to adequately explain why the peg passes through the square hole but not the round hole (Putnam, 1975). The accounts

offered by physics and chemistry fail as explanations because they hide, in a forest of irrelevant physical detail, the features most relevant to the phenomena being described. In the case of the peg, the relevant features are the geometric properties of the peg and board, and in the case of mental properties, the relevant features are those higher-order, functional properties mentioned in functionalist theories of mind. Since a reducing theory should adequately explain the phenomena of the reduced theory, it follows that mentalistic theories cannot be reduced to those of the physical sciences. And if so, then it would seem that the functional features definitive of mentality *merely* supervene on the underlying physical features.

Granted, the goal of explanation is to provide an understanding of the features being explained. So if the functional features that are essential to mentality are lost in the detail of physical descriptions, then physical descriptions would not count as adequate explanations of mentality, if they count as explanations at all. In this sense, mentalistic theories are irreducible to the theories of the physical sciences. However, if the point is merely *epistemic*, then this is not the sort of irreducibility that entails either NR_1 or NR_2 . Despite the plethora of irrelevant detail provided by the physical account of some mental property M, it might still be that there is something in common, at the physical level, to all of the events that realize M, and it might also be that this common feature is both necessary and sufficient for (and possibly even identical with) the functional property definitive of M. Thus, the fact that the descriptions offered by the physical sciences obscure the functional features that are crucial to mentality does not entail that these functional features merely supervene.

It is not enough that the physical sciences fail to capture what is essential to mentality in an explanatorily adequate way – e.g., in a way that facilitates understanding. The proponent of NR_1 and NR_2 needs to show that the physical sciences ignore those features *entirely*. But why should we believe this? As shown above, multiple realizability is not a good reason. Considerations of multiple realizability show that the events that realize some mental property M might differ greatly in physical (e.g., neural) type, and therefore no one of those types is necessary for the presence of M. However, multiple realizability does not show that the type of *interaction* that obtains at the physical level, in virtue of which an event manages to

realize M, might itself vary from case to case. So multiple realizability allows that all the events that realize M do so in virtue of standing in some physical relation R to other physical events. To show that relation R is something other than a physical relation, additional arguments are needed.

Van Gulick (1992) attempts to provide an additional argument by citing some important dimensions along which theoretical frameworks might differ. These include how the theory is used to guide one's interactions with the world, the goals to whose realization the theory can contribute, the role of indexical elements in fixing the content of the theory, and the individuating principles employed by the theory – especially, concerning how widely or narrowly types are individuated (pp. 166–168). Mentalistic theories often differ from the physical sciences in these respects. For example, externalist intuitions regarding mental content may lead us to classify two systems differently even though, when viewed from a neurological perspective, we would treat them as having all the same properties, and objects that have nothing in common when viewed from a molecular framework may be classified as similar in virtue of their shared capacity to produce phenomenally blue visual experiences in perceivers. On the basis of these differences, Van Gulick concludes that

... at least some aspects of our mentalistic framework are incommensurable with physical theory. There are concepts, such as those of property M and the functional relation R_m , that we can not capture within the conceptual resources of physical theory. (p. 166)

However, the fact that our mental and physical theories differ along the dimensions Van Gulick cites does not, by itself, establish either NR_1 or NR_2 . It must be shown not only that our mental and physical theories pick out different features of the world, but that this is the case *even in principle*. The individuating principles employed by *current* neurobiology allow that creatures might count as having all the same neural properties even though they differ mentally. But for all this shows, it is still possible that a different neurobiological framework might individuate events widely enough that there will never be a mental difference without some difference in neural properties. Thus, to establish NR_1 and NR_2 , it needs to

be shown that *no* theory that would count as a physical theory can capture the functional features definitive of mentality.

But if this can be shown, then in what sense is physicalism true? If the complete physicalistic story leaves out certain features that are essential to mentality, then the physical properties that are had would underdetermine the mental properties that are had. It may be that mental properties are underdetermined in just this way. Perhaps Nagel (e.g., 1974) and Searle (e.g., 1992) are right to think that mentality has an essentially subjective character which cannot be reduced to purely physical features. Or perhaps mentality is ineliminably normative in character. The attribution of mental content might be guided by rationality constraints (e.g., various principles of charity),⁹ and it may be that these normative factors are underdetermined by physical factors. In either case, the functional properties that ensure the presence of mentality could not be characterized purely physically. However, while these theories may be true, they are certainly not available to the *physicalist*. If the complete physicalist story fails to capture what is essential to mentality, then mentality would not be had solely in virtue of physical features.

To sum up: there seems to be no good reason to think that the functional features definitive of mentality merely supervene on physical features. An appeal to multiple realizability does not establish this conclusion, and an appeal to the autonomy of the mental establishes NR_1 and NR_2 only under a very strong construal – one which implies that the physical sciences fail to capture what is essential to mentality, even in principle. However, this strong sense of irreducibility would also be a reason for denying the belief in supervenience that underlies physicalism. So for the non-reductionist who wishes to be a physicalist, the dilemma presented at the end of section 3 remains a serious threat. Either we give up the belief in supervenience thesis S, which requires giving up physicalism, or we identify the functional features characteristic of mentality with physical features, which amounts to giving up non-reductionism.

5. CONCLUSION

Van Gulick contends that “those of us who want to reject reductionism but keep our materialist credentials can continue to ‘eat our

cake and have it, too'." (1992, p. 178). This is certainly an attractive option, but it ignores a crucial tension. There is no tension between supervenience and non-reductionism *per se*. S entails that there are nomologically true conditionals of the form 'P \rightarrow M,' linking physical properties to mental properties, but it does not guarantee nomologically true conditionals of the form 'M \rightarrow P'. Thus, S is compatible with NR₁.¹⁰ S would also be compatible with NR₂, since true biconditionals is a prerequisite for true identity claims.

Problems arise when S is conjoined with both non-reductionism and physicalism. The most common motivation for being a non-reductionist is the belief that something other than intrinsic physical type determines whether an internal event manages to instantiate some mental property. This something other is typically thought to be the functional role played by the internal event. But whatever it is, so long as it is something other, mental properties will not supervene on physical properties, unless this additional feature is either (i) a physical property or (ii) a non-physical property that merely supervenes on physical properties. Accepting option (i) amounts to giving up NR₁ and NR₂. And, as shown in section 4, accepting (ii) seems entirely unmotivated. If, on the other hand, we reject supervenience, we no longer have a theory that qualifies as "physicalistic."

These worries arise only for the physicalist who accepts NR₁ and NR₂, but as noted in section 4, there is another way to earn the title "non-reductionist." One can reject reductionism on *epistemic* rather than ontological grounds. The physical sciences, one might argue, do not adequately explain (i.e., in a way that enhances understanding) the functional features definitive of mentality. If this is right, then there is a sense in which mentalistic theories cannot be reduced to theories of the physical sciences, even though both NR₁ and NR₂ may be true.

Alternatively, one can take on the challenge of redefining 'physical' in such a way that mental properties may count as genuinely physical properties without having to supervene on the features mentioned by the physical sciences. These two options may be as close as we can get to "eating our cake and having it, too."

NOTES

¹ See, for example, Putnam (1967), Davidson (1970), Fodor (1974), Boyd (1980), and Van Gulick (1992).

² Robert Richardson (1979) convincingly argues that successful scientific reduction is often carried out in the absence of true biconditionals. He also notes that true biconditionals is something that Nagel himself did not require.

³ See Crane and Mellor (1990) for an excellent discussion of these and other attempts to set the boundaries of the physical.

⁴ The second modal operator ensures what Kim's "weak" supervenience (1984(a), p. 163) does not – namely, that physical indistinguishable creatures are also mentally indistinguishable, *even across possible worlds*.

⁵ Of course, not all physical relations are relevant to mentality. Existing in a world where the rings of Saturn have $x + 1$ rather than x ammonia molecules does not seem relevant; nor would existing in a world that has one more grain of sand on Malibu Beach. Since these physical features are irrelevant, they would be excluded from the supervenience base. And since many specific physical relations would be excluded, so too would the more general relational property – existing in a world that is physically indiscernible from the actual world.

These are physical properties that we would obviously want to exclude. In other cases, the decision is more difficult. Accepting S, however, does not require that we have actually made all the difficult decisions. Suffice it that there is some set of physical properties such that for any mental property M, there is at least one member of that set (whatever it might be) that ensures the presence of M. Deciding exactly which physical properties are included in that set, and then deciding which mental properties are ensured by the presence of each of those physical properties would be the task we face when arriving at a complete and precise physicalistic theory of mind on the basis of the general intuition expressed by S.

⁶ Functionalists of the "analytic" variety do insist on one type of conceptual connection. They believe that for any mental property M, there is a functional role F such that 'If x has an internal event that plays F, then x has M' is conceptually true. But, of course, this is very different from claiming that the physical (e.g., neural) events that might play role F are themselves conceptually tied to mental properties.

⁷ The word 'intrinsic' is misleading. The neural features of an event are not entirely intrinsic, since neural types are individuated at least partly in terms of the causal roles played by their instances. However, for the present discussion, this complication can be ignored. As far as U_1 and U_2 are concerned, "intrinsic" can include any property individuated narrowly enough so that its instances are not guaranteed to play the functional roles definitive of mentality.

⁸ To bring out more clearly the idea that an instance of a mental property is an event that plays the right functional role with respect to other events, the functionalist may wish to express functional property F as follows: having an event e which satisfies the functional schema ' $(\exists x_2)(\exists x_3) \dots (\exists x_n)C(e, x_2, x_3, \dots x_n)$.'

⁹ See, for example, Davidson (1975).

¹⁰ Kim notes that strong supervenience does give us true biconditionals if we assume that supervenience bases are closed under disjunction. For any mental property M, there will be some physical property which not only guarantees its presence, but is also guaranteed by its presence. This physical property is expressed

by a very long (possibly infinite) disjunction, the disjuncts of which denote each of the physical properties whose instances might realize M. (See, for example, Kim's 1990, pp. 19–20.)

The typical non-reductionist response is that the disjuncts invoked will be far too heterogeneous to express genuine laws. Thus, the biconditionals in which they appear will be unsuited for theoretical reduction. (See Fodor (1974) and Kim (1992).)

REFERENCES

- Boyd, R. 1980, 'Materialism Without Reductionism: What Physicalism Does Not Entail', in N. Block (ed.), *Readings in Philosophy of Psychology*, vol. 1 (Cambridge, Mass.: Harvard University Press), pp. 67–106.
- Crane, T. and Mellor, A. 1990, 'There is No Question of Physicalism', *Mind* 99, pp. 185–206.
- Davidson, D. 1970, 'Mental Events', in L. Foster and J. W. Swanson (eds.), *Experience and Theory* (Amherst: University of Massachusetts Press), pp. 79–101. Reprinted in Davidson (1980), *Essays on Actions and Events* (Oxford: Oxford University Press), pp. 207–225.
- Davidson, D. 1975, 'Thought and Talk', in S. Guttenplan (ed.), *Mind and Language* (Oxford: Oxford University Press), pp. 7–23. Reprinted in Davidson (1984) *Truth and Interpretation* (Oxford: Clarendon Press).
- Fodor, J. 1974, 'Special Sciences, or The Disunity of Science as a Working Hypothesis', *Synthese* 28, pp. 97–115.
- Jack, A. 1994, 'Materialism and Supervenience' *Australasian Journal of Philosophy* 72, pp. 426–444.
- Kim, J. 1984(a), 'Concepts of Supervenience', *Philosophy and Phenomenological Research* 45, pp. 153–176.
- Kim, J. 1984(b), 'Epiphenomenal and Supervenient Causation', *Midwest Studies in Philosophy* 9, pp. 257–270.
- Kim, J. 1987, "'Strong" and "Global" Supervenience Revisited', *Philosophy and Phenomenological Research* 48, pp. 315–326.
- Kim, J. 1989, 'The Myth of Nonreductive Materialism', *Proceedings and Addresses of the American Philosophical Association* 63, pp. 31–47.
- Kim, J. 1990, 'Supervenience as a Philosophical Concept', *Metaphilosophy* 21, pp. 1–27.
- Kim, J. 1992, 'Multiple Realization and the Metaphysics of Reduction', *Philosophy and Phenomenological Research* 52, pp. 1–26.
- Kim, J. 1993(a), 'The Nonreductivist's Troubles with Mental Causation', in J. Heil and A. Mele (eds.), *Mental Causation* (Oxford: Oxford University Press), pp. 189–210.
- Kim, J. 1993(b), 'Postscripts on Supervenience', in *Supervenience and Mind: Selected Philosophical Essays* (Cambridge: Cambridge University Press), pp. 161–171.
- Moser, P. 1992, 'Physicalism and Global Supervenience', *The Southern Journal of Philosophy* 30, pp. 71–82.

- Nagel, T. 1974, 'What Is It Like To Be a Bat?', *Philosophical Review* 83, pp. 435–450.
- Putnam, H. 1967, 'Psychological Predicates', in Capitan and Merrill (eds.), *Art, Mind and Religion* (Pittsburgh: University of Pittsburgh Press), pp. 37–48. Reprinted as 'The Nature of Mental States', in *Mind, Language and Reality: Philosophical Papers II* (Cambridge: Cambridge University Press, 1975), pp. 429–440.
- Putnam, H. 1975, 'Philosophy and Our Mental Life', in *Mind, Language, and Reality: Philosophical Papers II* (Cambridge: Cambridge University Press), pp. 291–303.
- Richardson, R. 1979, 'Functionalism and Reductionism', *Philosophy of Science* 46, pp. 533–558.
- Searle, J. 1992, *The Rediscovery of the Mind* (Cambridge, MA: MIT Press).
- Teller, P. 1986, 'Relational Holism and Quantum Mechanics', *The British Journal for the Philosophy of Science* 37, pp. 71–81.
- Van Gulick, R. 1992, 'Nonreductive Materialism and the Nature of Intertheoretic Constraint', in A. Beckermann, H. Flohr and J. Kim (eds.), *Emergence or Reduction?: Essays on the Prospects of Nonreductive Materialism* (Berlin: Walter de Gruyter), pp. 157–178.

Department of Philosophy
San Diego State University
San Diego, CA 92182
USA