

## HOW SUPERDUPER DOES A PHYSICALIST SUPERVENIENCE NEED TO BE?

BY JESSICA WILSON

It has recently been made clear that the standard formulations of the supervenience relation – weak, global and strong, among others – are too programmatic to be the full answer to a physicalist's prayers.<sup>1</sup> All the standard formulations present the supervenience of one set of properties on another in terms of property correlations, without placing any constraints on the dependency relation effecting these correlations. It turns out that such abstract characterizations do not ensure that properties supervening upon (what are assumed to be) physicalistically acceptable base properties are themselves physicalistically acceptable. Most disastrously, the standard formulations of supervenience turn out one and all to permit supervenient properties to be emergent in a way at odds with materialism. What physicalism needs to support a materialist metaphysics is 'superdupervenience' – supervenience that guarantees that supervenient properties are 'nothing over and above' their physicalistically acceptable base properties. Terence Horgan (p. 563) has suggested that the following constraint will do the job:

*Horgan's constraint.* Any genuinely physicalist metaphysics should countenance ontological inter-level supervenience relations only if they are robustly explainable in a physicalistically acceptable way.

He uses the qualifiers 'ontological' and 'robustly explainable' to stop the explanation from showing (albeit in a physicalistically acceptable way) that the supervenience relation reflects a merely conceptual or semantic constraint on a given domain of discourse: 'supervenience is *ontological* if it is

<sup>1</sup> See, e.g., S. Schiffer, *Remnants of Meaning* (MIT Press, 1987); J. Kim, 'Supervenience as a Philosophical Concept', *Metaphilosophy*, 21 (1990), pp. 1–27; J. Heil, *The Nature of True Minds* (Cambridge UP, 1992); T. Horgan, 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World', *Mind*, 102 (1993), pp. 555–85.

an objective relation between lower-level properties and facts and genuine, objective, higher-level properties and facts ... supervenience for a given mode of discourse is *robustly explainable* if it is explainable as ontological'.

A requirement of robust explanation may seem like a natural way for physicalists to go about blocking the possibility of emergent supervenient properties. After all, the absence of robust inter-level explanations fuelled turn-of-the-century emergentist claims that chemical and biological properties were emergent; and the advent of explanations for these properties, in terms of base properties that were presumed physicalistically acceptable, coincided with the decline of emergentism and the rise to prominence of various physicalist accounts of empirical properties. But if robust explanation, interpreted as displaying how the characteristic features of supervenient and base properties match up, is required to block the possibility of emergent properties, then physicalism's apparent triumphs over emergentism may be at an end. For, as Horgan points out, the presence of various 'explanatory gaps' between mental and physical properties makes it unlikely that Horgan's constraint can be met for the case of mental properties. The prospect of supervenience's forming the basis for a specifically physicalist bridge between the physical and the mental might then appear dim.

Against such pessimism, I shall argue that, for an important class of cases, those in which supervenient and base properties are instantiated in the same individuals, robust explanation is neither sufficient nor necessary to ensure that properties supervening on presumed physicalistically acceptable base properties are themselves physicalistically acceptable. Most varieties of supervenience are addressed to properties in this class. Strong supervenience, for example, is formulated with respect to sets of properties  $A$  and  $B$ , elements of which are concurrently instantiated in individuals drawn from a domain  $D$ , as follows:  $A$  strongly supervenes on  $B$  iff  $\Box(\forall x \in D)(\forall a \in A)(x \text{ has } a \supset (\exists b \in B)(x \text{ has } b \text{ and } \Box(\forall y \in D)(y \text{ has } b \supset y \text{ has } a)))$ . The class of cases in which supervenient and base properties are instantiated in the same individuals is important, for, as we shall see, it is in these cases that the threat of systematic causal overdetermination is most worrisome.

To begin with, I shall show that Horgan's paradigm of such explanation (one of the class of cases at issue) is compatible with the supervenient property's being emergent. I shall go on to argue that this and other unacceptable possibilities may be ruled out, even in the absence of robust explanation, by means of a metaphysical constraint on the supervenience relation. For empirical properties, it suffices to ensure the physicalistic acceptability of properties supervening on physicalistically acceptable base properties (where the properties are instantiated in the same individuals) that the sets of causal powers associated with supervenient and base

properties stand in an appropriately formulated *internal* relation. In particular, each individual causal power in the set associated with a given supervenient property must be numerically identical with a causal power in the set associated with its base property. Satisfying this constraint on causal powers is all that is needed to render supervenience superduper.

## I

Investigations into the metaphysics of empirical properties usually take as a starting-point a familiar ontological picture, inspired by the special sciences, according to which empirical reality exhibits a layered hierarchical structure. For example, Jaegwon Kim characterizes the layered framework as

a hierarchically stratified structure of 'levels' or 'orders' of entities and their characteristic properties. It is generally thought that there is a bottom level, one consisting of whatever microphysics is going to tell us are the most basic physical particles out of which all matter is composed.... As we ascend to higher levels, we find structures that are made up of entities belonging to the lower levels, and, moreover, the entities at any given level are thought to be characterized by a set of properties distinctive of that level.<sup>2</sup>

The picture is usually presented as giving rise to two questions: first, what is the nature of the relations holding between entities at different levels? And second, what is the nature of the relations holding between properties at different levels? It is the second question which is, as Kim says (p. 191), 'the crucial question answers to which have defined various currently contested positions on certain metaphysical and methodological issues including, most notably, the mind-body problem'.

While there is much to recommend taking this ontological picture as an appropriate starting-point for investigations into empirical properties, there is a sense in which the picture is misleading. For it can lead one to think that the only inter-level relation between properties at issue is that between properties instantiated in *different* entities (since, as the picture is usually drawn, different entities exist at different levels). It is indeed crucial to investigate the relation holding between, for example, the property of having a certain charge, instantiated in electrons and protons, and the bonding properties that are instantiated in atoms made up of these constituents. It is also crucial, however, to investigate the relation holding between apparently different properties that are instantiated in the *same* entities. For example, we

<sup>2</sup> J. Kim, 'The Non-Reductivist's Troubles with Mental Causation', in J. Heil and A. Mele (eds), *Mental Causation* (Oxford: Clarendon Press, 1993), pp. 189–210, at p. 190.

want to understand how the property of fragility, instantiated in some quantity of stuff, is related to the property of having a certain molecular structure, instantiated in the same quantity of stuff. And we are particularly interested in understanding how the mental and physical properties of persons are related. It is this latter sort of ‘inter-level’ relation between properties on which I shall focus here.

Supervenience has often been suggested as a promising way for physicalists to characterize how properties at different levels may be related. A physicalist supervenience – roughly, that there is (or could be) no change in properties without a change in physical properties – has seemed an apt expression of the physicalist thesis that the physical facts, appropriately defined, determine all the facts. But as Horgan and others have pointed out, the historical uses of supervenience may have thrown cold water on these prospects. The concept of supervenience, if not the word, seems to have been first invoked by the moral *non*-naturalist G.E. Moore; and R.M. Hare, who did use the term, did so in support of an irrealist position on supervenient moral properties.<sup>3</sup> Moreover, British emergentists (Samuel Alexander, C.D. Broad, C. Lloyd Morgan) used the term in formulating their theses, and, as Horgan notes, these accounts (some version of which was meant to contrast with physicalism) were compatible with emergent properties’ being ‘physically’ supervenient, in the sense that there could be no change in emergent properties without a change in physical properties, on the ‘basal’ conditions from which they emerged.<sup>4</sup> Historical formulations of supervenience, then, were compatible with supervenient properties’ being non-natural, unreal or emergent. And it is not hard to see that the standard contemporary formulations do not rule out these untoward possibilities. Like their predecessors, contemporary formulations characterize supervenience by means of property correlations, and they differ from earlier formulations only in so far as they require that the correlations hold in a more or less wide range of possible worlds, or that the base properties span a more or less wide region of space-time. Since contemporary formulations remain silent on the details of the dependency relation implementing these correlations, they are compatible with this relation’s being such that properties supervening on physicalistically acceptable base properties could none the less be emergent, or otherwise physicalistically unacceptable. Neither historical nor contemporary formulations of supervenience are, as they stand, up to the task of grounding a (realist) physicalist metaphysics.

<sup>3</sup> G.E. Moore, *Principia Ethica* (Cambridge UP, 1903); R.M. Hare, *The Language of Morals* (Oxford UP, 1952).

<sup>4</sup> S. Alexander, *Space, Time and Deity* (London: Macmillan, 1920); C.D. Broad, *The Mind and its Place in Nature* (London: Routledge & Kegan Paul, 1925); C.L. Morgan, *Emergent Evolution* (London: Williams & Norgate, 1923).

According to Horgan, what needs to be done in order to ensure that supervenience preserves physicalistic acceptability is that Horgan's constraint must be imposed: any posit of supervenience must be accompanied by a robust, physicalistically acceptable explanation of the supervenience connection at issue. What would such an explanation look like? Horgan offers the following three questions as constitutive of the task:

1. *Target question.* What facts that need explaining are such that explaining them in a physicalistically acceptable way would constitute a physicalistically acceptable explanation of a given inter-level relation?
2. *Standpoint question.* What sort of facts, over and above physical facts and laws, could (in combination with physical facts and laws) yield a physicalistically acceptable explanation of the inter-level relation?
3. *Resource question.* Do there exist adequate explanatory resources to provide such an explanation?

He goes on (p. 579) to discuss liquidity as an example of a supervenient property susceptible to a successful robust explanation in terms of physicalistically acceptable base properties:

Explaining why liquidity supervenes on certain microphysical properties is essentially a matter of explaining why any quantity of stuff with these microphysical properties will exhibit these macro-features [tendency to flow, to assume shape of vessel that contains it, etc.]. (As regards the target question, this suffices to explain the supervenience of liquidity because those macro-features are *definitive* of liquidity. As regards the standpoint problem, it seems explanatorily kosher to assume a 'connecting principle' linking the macro-features to liquidity, precisely because those features *are* definitive; the connecting principle expresses a fact about what liquidity is.)

This case is, of course, one of the class of cases for which the relation in question holds between properties instantiated in the same individual: there is a single 'quantity of stuff' which instantiates both certain microphysical properties and the higher-level property of liquidity. Here the target facts needing explanation are (the exhibition of) certain macro-features associated with liquidity. The resource question is answered affirmatively, for explanations exist which make it intelligible how any quantity of stuff instantiating such and such microphysical properties will exhibit these macro-features. And the answer to the standpoint question tells us that these explanations are good enough, since we can get to liquidity from the microphysical explanations of the macro-features in question, with the help of a single definitional fact connecting the macro-features with liquidity. Since physicalistically acceptable facts, when combined with definitional facts, remain physicalistically acceptable, the explanation as a whole is physicalistically acceptable.

Horgan's example makes clear that, on his account, there will not be a *general* formulation of a 'physicalistically acceptable' supervenience relation: each case of supervenience will need to be considered separately to see if we have the resources to explain what needs to be explained in a physicalistically acceptable way. One case for which he doubts that we have the resources required is the case of mental properties. He offers as illustration a candidate configuration of the target, standpoint and resource questions which aims to establish the supervenience of mental properties on (presumed physicalistically acceptable) behavioural properties. The facts needing explanation are taken to be the generalizations of common-sense psychology. As in the case of liquidity, these generalizations are plausibly definitive of (at least a range of) mental properties, so that explaining these features in a physicalistically acceptable fashion, in combination with such definitional facts, would explain satisfactorily the mental properties in question. The problem in this case, as Horgan sees it (p. 580), is that we lack the resources to explain the target facts thus in terms of the physicalistically acceptable 'base' facts: since, as according to Quine,<sup>5</sup> there will evidently be vastly many incompatible ways of assigning propositional attitudes to people over the course of their lifetimes, all of which would be compatible with their obeying the generalizations of common-sense psychology, 'the constraint does not suffice to yield *determinate* supervenience connections between physical properties and facts and mental ones'.

Even if one has faith that a gap based on indeterminacy can be bridged, other explanatory gaps lie in wait. There is, for example, the is/ought gap. Naturalizing projects in meta-ethics have had considerable difficulties in providing tractable specifications of normative properties or facts in physicalistically acceptable terms, and if, as Tyler Burge claims,<sup>6</sup> there is a normative element in intentional content, these difficulties will attach to the present project of robustly explaining the supervenience of mental on physical properties. And the case of *qualia* opens up an explanatory gap between qualitative and non-qualitative (presumably physical) properties, which has by now become a standard refrain:

There are also the 'phenomenal' or 'what-it's-like' mental properties to deal with, the so-called '*qualia*'. *Prima facie*, it is enormously hard to see how one could possibly explain why any particular physical or neurobiological property always gets co-instantiated with (or why it *necessarily* always gets co-instantiated with) a particular phenomenal property – or with any phenomenal property at all.<sup>7</sup>

<sup>5</sup> W.V.O. Quine, *Word and Object* (MIT Press, 1960).

<sup>6</sup> T. Burge, 'Individualism and Psychology', *The Philosophical Review*, 95 (1986), pp. 3–45.

<sup>7</sup> J. Levine, 'Materialism and *Qualia*: the Explanatory Gap', *Pacific Philosophical Quarterly*, 64 (1993), pp. 354–61, at p. 354. See also D. Chalmers, *The Conscious Mind* (Oxford UP, 1996).

If these philosophers are right, the chances that physicalists will be able to satisfy Horgan's constraint for the case of mental properties are slim. Just to stack the deck, other features characteristic of mentality – the subjectivity and the unity of conscious experience, for example – promise to give rise to still more explanatory gaps. I, for one, am prepared to agree not only that it is currently more or less opaque how the various properties distinctive of mentality might be explained in terms of physicalistically acceptable properties, but also that, given the seemingly *transcendent* nature of certain features of mental properties, compared with features characteristic of physical properties, it is likely that there might never be any such explanations, even in principle. In the light of this, are the prospects of a superdupervenience that works for the case of mental properties hopeless?

## II

I do not think the prospects are as bad as Horgan and others claim. I shall show this in two steps. First, I shall show that the robust explanation which Horgan offers as a paradigm case of an explanation grounding a supervenience claim, liquidity, is compatible with liquidity's being emergent in a way at odds with physicalism. This shows that the requirement of robust explanation, construed as a demand for intelligibility, is not sufficient for the task of formulating a physicalistically acceptable supervenience. Second, I shall argue that considerations of what is really at issue between emergentists and physicalists provide reason to think that such a requirement is not necessary to the task either.

Horgan claimed that liquidity was a case of superdupervenience, on the grounds that, first, certain macro-features were definitive of liquidity (tendency to flow, tendency to conform to the shape of an enclosing container, etc.); and second, we have in hand a physicalistically acceptable explanation of why any quantity of stuff with certain microphysical properties will exhibit these macro-features. Now, is this explanation sufficient to *guarantee* that liquidity is a physicalistically acceptable property?

The answer is no. For surely one 'physicalistically acceptable explanation of why any quantity of stuff with certain micro-properties will, in proper circumstances, exhibit certain macro-features' would be for the micro-properties to *cause* the macro-features. John Searle<sup>8</sup> is happily explicit on the point: 'There are many examples in nature where a higher-level feature of a system is caused by lower-level elements of that system, even though the

<sup>8</sup>J. Searle, 'The Problem of Consciousness', in M. Michael and J. O'Leary-Hawthorne (eds), *Philosophy in Mind* (Amsterdam: Kluwer, 1994), pp. 174–90, at p. 181.

feature is a feature of the system made up of those elements. Think of the liquidity of water or the transparency of glass or the solidity of a table, for example.’ Searle’s statement is ambiguous between the two inter-level relations between properties mentioned earlier; whether he intends to suggest that the relation we are currently investigating is causal would depend on whether the property (instantiation) taken to cause the higher-level property (instantiation) was a microstructural property of the system. Whichever relation Searle has in mind, however, the possibility remains that, for all Horgan’s constraint tells us, a given higher-level property, instantiated in some entity, might be ‘robustly explained’ by reference to the property’s being caused by a lower-level microstructural property of that entity. For those who think that supervenience cannot be causation, since causation is a diachronic relation and supervenience is not, we can note that there is, none the less, as Stephen Yablo points out, a ‘subtle interpretative question’ about supervenience:

On the *emergence* interpretation [of supervenience of mental on physical properties], a thing’s physical properties are metaphysically prior to its mental properties and bring them into being. To caricature emergentism just slightly, supervenience is a kind of ‘supercausation’ which improves on the original in that supercauses act *immediately* and metaphysically *guarantee* their supereffects.<sup>9</sup>

Why is the possibility that supervenience might be a causal or supercausal relation a problem for physicalists? Yablo’s reference to supercausation as ‘emergent’ provides a clue, but requires some filling in. Here it helps to see how the British emergentist C. Lloyd Morgan (pp. 15–16) thought about emergent properties:

I speak of events at any given level of the pyramid of emergent evolution as ‘involving’ concurrent events at lower levels. Now what emerges at any given level affords an instance of what I speak of as a new kind of relatedness of which there are no instances at lower levels.... This we must accept ‘with natural piety’ as Mr Alexander puts it.... But when some new kind of relatedness is supervenient ... the way in which the physical events which are involved run their course is different in virtue of its presence.

In this passage Morgan is making two claims about emergent properties. The first is that no explanation should be expected regarding the appearance of such properties – they must be accepted ‘with natural piety’. The second is that emergent properties are efficacious in some way different from the lower-level properties from which they emerge. Now Horgan’s constraint can be seen as an attempt to ensure the physicalistic respectability of supervenient properties by requiring that an explanation be in place, so

<sup>9</sup> S. Yablo, ‘Mental Causation’, *Philosophical Review*, 101 (1992), pp. 245–80, at pp. 256–7.



blocking the application of Morgan's first claim to supervenient properties. But it is Morgan's second claim, that emergent properties enter into causing different effects, that is, have causal powers different from those of their base properties, which is, at least with respect to higher- and lower-level properties instantiated in the same individual, most crucial to distinguishing physicalism from emergentism.

We can see this by considering the different sorts of responses emergentists and physicalists give to the so-called 'problem of mental causation'. A weak form of what Kim calls 'Alexander's dictum' – that a real empirical property has causal powers – has for some time now been common ground in the present ontological debate. (This is not to say that empirical properties may be exhaustively analysed in terms of causal powers, but only that associated with every real empirical property is some set of causal powers.) So emergentists and physicalists who allow that (some) higher-level properties are real are committed to these properties' having causal powers. It is this commitment which leads to the problem of mental causation: given that mental properties have powers to cause various physical effects, and given that the physical is causally closed (so that every physical effect may be taken to have a purely physical cause), the physical effects of mental causes appear to be systematically overdetermined. The emergentist response to the possibility of causal overdetermination is to deny that the physical is causally closed – in other words, to deny that the physical effects of mental causes also have physical causes – and to assert that mental properties have causal powers not possessed by any physicalistically acceptable base properties. Physicalists committed to the efficacy of mental properties cannot take this route and allow that mental properties have any causal powers that are different from those of their physicalistically acceptable base properties, for this violates the physicalist thesis that mental properties are 'nothing over and above' their base properties. Nor can they allow that mental properties have the *same*, but still numerically distinct, causal powers as their base properties, for this leads to overdetermination. The remaining option is to hold that each of the causal powers associated with a given mental property is numerically identical with a causal power of its physicalistically acceptable base property. (As we shall see, this restriction does not entail that mental properties are identical with their base properties.)

This contrast between physicalist and emergentist treatments of higher-level properties is not limited to the case of mental properties. Although, in this case, the presumption of the causal closure of the physical makes for a quick route to the threat of causal overdetermination, it is not hard to see that the threat arises whenever there is a live possibility that two properties, instantiated in the same individual, may give rise to the same effects. To

deal with this threat of a generalized ‘problem of higher-level causation’, physicalists, in contrast with emergentists, will need to treat higher-level properties generally as not having any individual causal powers that are numerically distinct from some causal power of their base properties.

From the perspective of this generalized threat, we can see what was really at issue in the turn-of-the-century dispute between emergentists and physicalists over the status of chemical and biological properties. True enough, emergentists initially claimed that these properties were emergent, and they gave up these claims when explanations for these properties in terms of physicalistically acceptable properties came on the scene. But in the light of the generalized threat of causal overdetermination, it is clear that not just any explanation would do: an explanation would presumably encourage emergentists to retract the claim that a given higher-level property was emergent only if the explanation served to show that the property did not, after all, have any causal powers different from those associated with its base property. As such, it is not explanation *per se*, or the lack of it, that is crucial to distinguishing physicalistically acceptable properties from emergent properties. What *is* crucial is whether or not the following condition holds:

*Condition on causal powers (CCP).* Each individual causal power associated with a supervenient property is numerically identical with a causal power associated with its base property.

If a supervenient property has more than one base property, CCP must hold for each of them.

To return, then, to the case of liquidity: does the sort of ‘robust explanation’ provided by Horgan *entail* CCP? Since this explanation is compatible with liquidity’s being caused or supercaused by various lower-level properties, the answer is no. The reason is simple: effects (as property instantiations) generally (and at least sometimes) have causal powers that are distinct from those of their causes (as property instantiations). Effects are generally (and at least sometimes) ‘something over and above’ their causes. Causal relations do not guarantee the satisfaction of CCP, so a supervenience meeting Horgan’s constraint – one that is robustly explainable in a physicalistically acceptable way – is not superduper after all.

### III

Before laying out what sort of supervenience is superduper, I offer a few remarks about Horgan’s constraint. Although the above discussion shows

that robust explanation is not sufficient for the task of carving out an acceptable physicalist supervenience, the possibility remains that Horgan's constraint might be a necessary condition. One reason not to think so has to do with its being inappropriate to impose a distinctly epistemic condition on a decidedly metaphysical task. To the point here are some remarks of Kim's, which he makes in discussing LePore and Loewer's analysis of the realization relation, where property  $P$  realizes property  $M$  if (1)  $P \supset M$  with nomological necessity, and (2)  $P$  explains  $M$ . In discussing condition (2), Kim (p. 193) says

Consider ... their explanatory requirement. I believe the idea behind this requirement is correct, but we should look for an objective metaphysical relation between  $P$  and  $M$ , not an essentially epistemic relation like explanation; that is, we should view the explanatory relation between the two properties as being supported by a metaphysical realization relation. Here I am taking a realist attitude about explanation: if  $P$  explains  $M$ , that is so because some objective metaphysical relation holds between  $P$  and  $M$ .

I agree with Kim that, in the present context, explanations are best seen as markers of metaphysical relations. Moreover, there are three points to consider: first, if what we are after is a metaphysical relation between properties, and if there happen to be *metaphysical* markers of this relation, then why require additional markers in the form of explanations? I think there are such metaphysical markers for the cases at hand, which I shall discuss shortly. Second, the metaphysical relation at issue is supposed to serve as the partial ground of a *unified* physicalist metaphysics of empirical properties, but there is nothing very unified about the sort of explanations that would be given as meeting Horgan's constraint. To avoid such a piecemeal approach, it seems preferable to avoid if at all possible having to rely on explanations as markers of the metaphysical relation in question. Third, given that explanation is epistemic, and given that our epistemic access is a matter both of our own capacities and of the complexity of the *explanandum*, robust explanations of many interesting higher-level properties may be out of reach for creatures like us. But it seems wrong to take such contingencies to tell against the possibility of a unified physicalist metaphysics, especially if there are independent reasons – for example, the threat of causal overdetermination – for thinking that the metaphysical relation is in place. As Kim notes, if  $P$  explains  $M$ , this is plausibly because some metaphysical relation holds between  $P$  and  $M$ . But surely the entailment need not go the other way: even if an objective metaphysical relation holds between  $P$  and  $M$ ,  $P$  might not explain  $M$ . Why, then, take explanation as a necessary precursor to metaphysical deliberation?

One reason for doing so might be if the only way to get a handle on the properties involved were by reference to their definitive or characteristic features. As we have seen, Horgan takes this approach in investigating the property of liquidity and mental properties; in both cases the viability of the supervenience claims turned on whether certain definitive features of the putatively supervenient property could be explained in terms of certain definitive features of the base property. Perhaps if the only way in which properties could be picked out were by reference to their definitive features, a condition on explanation (connecting up these definitive features) would be necessary in order to show that one property superdupervenied on another.

But reference to definitive features of a property is not the only way one might go about picking out an empirical property. Given that the relevant point of contention between physicalism and emergentism (again with respect to properties instantiated in the same individuals) is that physicalism asserts and emergentism denies CCP, there is an obvious alternative: pick out properties by reference to their causal powers. (Of course, functionalists will think that possession of certain causal powers *is* definitive of some or all properties. But the present suggestion is meant to be contrasted with the suggestion that properties should be picked out via reference to characteristic features like those usually seen as admitting of an explanatory gap, e.g., consciousness, intentionality, normative content, which are not usually cashed in terms of specific causal powers.) Should we do this, we can reconceive the task of determining whether a given higher-level property supervenes on a lower-level base property. In Horgan's terms, the task will involve determining (as target question) what causal powers are associated with the higher-level property, and determining (as resource question) whether we are capable of ascertaining whether each causal power possessed by the higher-level property is identical with a causal power of the candidate base property. If this task could be accomplished with affirmative results, then (given that the requirement which is needed to ensure that a given property is not emergent is built into the process) CCP *would* be entailed.

A physicalist supervenience that ensures the satisfaction of CCP blocks not only the possibility of the supervenient property's being emergent, but also (in conjunction with Alexander's dictum) seems likely to block the other ways in which in the past it has been argued that a supervenient property might be physicalistically unacceptable. We saw earlier that an unconstrained supervenience relation was compatible with supervenient properties' being either non-natural (Moore) or unreal (Hare). Now unless a better suggestion comes along, it seems plausible to assume that if a property

is non-natural or unreal, it must have some non-natural causal powers, or no causal powers at all. But these possibilities are excluded if a supervenient property cannot have any causal powers distinct from those of its base property. For by Alexander's dictum the supervenient property has some causal powers (and so is not unreal), but CCP ensures that (since the base property is assumed physicalistically acceptable, and so without non-natural causal powers) none of these causal powers is non-natural.

Conceiving of a physicalist supervenience in terms of causal powers would do more than allow physicalists to determine whether a given property superduperved on another. It would also provide a method for non-reductivists to establish that a given supervenient property is distinct from its base property, by showing that the base property has (one or more) causal powers *different* from those of the supervenient property. In this case the set of causal powers associated with the supervenient property would be a proper subset of the set of causal powers associated with the base property, thus providing a clear-cut account (which has too often been lacking) of how a higher-level (say, mental) property could be distinct from, and yet 'nothing over and above', a lower-level physicalistically acceptable property.<sup>10</sup> Attention to causal powers would also go some distance towards enabling physicalists to establish that a given property is multiply realizable, by showing that the causal powers associated with a given supervenient property are subsets, proper or improper, of the causal powers of properties in *more than one* supervenience base set. If we are looking for markers which can do some metaphysical work, causal powers seem to be just the ticket.

#### IV

CCP requires the causal powers of the supervenient property to be internally related, in a particular way, to the causal powers of its base properties. Following Armstrong's discussion, an *internal* relation is one that is dictated solely by the nature of the *relata*, and is such that, given certain entities with certain natures, the relation must hold between the entities.<sup>11</sup> Self-identity is the paradigmatic case of an internal relation; given any existent, then the simple fact of its existing guarantees that the relation holds. In contrast, spatio-temporal relations and causal relations are paradigmatically external relations. Given my nature and the nature of my

<sup>10</sup> This observation is due to Michael Watkins, who used it (in John Heil's 1996 NEH seminar in the Metaphysics of Mind) in service of a non-reductive physicalist solution to the problem of mental causation.

<sup>11</sup> D.M. Armstrong, *Universals: an Opinionated Introduction* (Boulder: Westview, 1989).

bookcase, I might or might not stand in the relation of being six feet from my bookcase; given the nature of these chemicals and the nature of that explosion, the first might or might not have caused the second. CCP, in requiring that each *individual* causal power of a supervenient property must be identical with a causal power of its base property, may be seen as requiring that the two *sets* of causal powers associated with the supervenient and base properties must be appropriately internally related. This requirement would be satisfied if the two properties were identical; hence reductive physicalist accounts guarantee the satisfaction of CCP in the most straightforward of fashions. But, as previously noted, the requirement can be satisfied by internal relations that are weaker than identity – if, for example, the set of causal powers associated with the supervenient property were a proper subset of the set associated with the base property.

From this perspective, the reason why standard accounts of supervenience are compatible with a supervenient property's being physicalistically unacceptable is that they are compatible (no matter what constraints they place on the property correlations in question) with an external relation's holding between the sets of causal powers associated with the supervenient and base properties. And the internal/external distinction explains why Horgan's constraint ultimately did not help with such matters, since robust explanation does not guarantee that the relation between the causal powers of these properties is *not* external. On the other hand, it seems clear that a supervenience relation which fulfils CCP avoids the unacceptable scenarios precisely because requiring the sets of causal powers to be internally related leaves no room for a supervenient property to be other than physicalistically acceptable, assuming the acceptability of the base property.

I now want to run through three non-reductive physicalist accounts of the supervenience (or realization) relation, and I shall show that the relation proposed in each case is one which implicitly or explicitly imposes an appropriate sort of internal relation (that is, one that guarantees fulfilment of CCP) on the sets of causal powers associated with the properties involved.

I shall start with Yablo's account of a physicalist inter-level dependency relation as a species of determinable/determinate relation, such as holds between, for example, red (as determinable) and scarlet (as determinate of red). Yablo sees the determinable/determinate relation between properties as an extension of the identity relation between properties, as expressed via the indiscernibility of identicals:

*P* is identical with *Q* only if necessarily, for all *x*, *x* has *P* iff *x* has *Q*.

An extension of the identity relation is the determination relation:

$P$  determines  $Q$  only if (i) necessarily, for all  $x$ , if  $x$  has  $P$  then  $x$  has  $Q$ ; and (ii) possibly, for some  $x$ ,  $x$  has  $Q$  but lacks  $P$ .

Traditionally, doctrines of identity and determination took the necessity involved in both cases to be conceptual necessity. But contemporary analysis being more sensitive to opaque contexts and the *a posteriori* necessities for which they make room, Yablo sees it as appropriate to take the necessity in question to be metaphysical, rather than conceptual.

Yablo suggests that we would do well to conceive of the relation between higher-level and lower-level properties as a variety of the metaphysically construed determinable/determinate relation, since this relation captures two desiderata commonly associated with physicalist accounts of mental properties. First, taking mental properties to be determinables of physical properties is compatible with the multiple realizability of the mental, for determinables are, commonly, multiply realized in their determinates. Second, this conception of the mental allows mental properties to be efficacious, without this efficacy introducing overdetermination, for determinables may be efficacious without being causal competitors with their determinates.

It is not hard to see that Yablo's suggestion implicitly trades upon the fact that the determinable/determination relation is an *internal* relation. The point comes out clearly in his discussion (p. 259) of how determinables do not causally compete with their determinates:

we know that [determinables and determinates] are not causal rivals. This kind of position is of course familiar from other contexts. Take for example the claim that a space completely filled by one object can contain no other. Then are even the object's *parts* crowded out? No. In this competition wholes and parts are not on opposing teams.

Determinables and their determinates, like objects and their parts, are guaranteed to be on the same team because they are internally related.

Yablo expected the suggestion, that mental properties stand to their physical realizations in the same relation as colours bear to their shades, to be met with some incredulity. One way to make his conjecture more plausible is to put the point in terms of the causal powers of the properties involved. A determinable property, like being red, has a variety of causal powers, including, say, the power to induce Sophie the pigeon to peck at a decent-sized patch with this property if it is placed in front of her. Now Sophie will, in the process of pecking at a red patch, also be pecking at a red patch of a particular shade, say scarlet. But even though the patch's being

scarlet also has the causal power to cause Sophie to peck at it (since to be scarlet just is to be red, in a particular way), there is no danger of these two properties' inducing any causal overdetermination with respect to Sophie's pecking. And the reason *why* this is so – the reason why 'determinables and their determinates are not causal rivals' – is because it is plausible, in the case of determinables and determinates, that each causal power of the determinable is identical with a causal power of its determinate. Now Alice, Sophie's cousin, *only* pecks at scarlet patches, so the property of being scarlet, which is a determinate of the property of being red, evidently has *more* causal powers than the property of being red. This is enough to show that being red is distinct from being scarlet (or being any of red's determinates), since the set of causal powers associated with being red is a proper subset of each of the sets of causal powers associated with its determinate properties. Moreover, the fact that the set of causal powers associated with being red is a subset of each of the sets associated with its determinates indicates that being red is, like many higher-level properties, multiply realizable.

Next I shall consider Philip Pettit's definition of physicalism. According to this definition, the physical facts which determine all other empirical facts will be those that appear in a true completed microphysics: 'Once the microphysical conditions and the microphysical laws have been fixed, then all the crucial features of a world like ours will have been fixed; *viz.*, all the laws that obtain at the world, all the conditions – all the initial conditions – that engage these laws and all the things that happen in accordance with the laws'.<sup>12</sup> Tim Crane objected to Pettit's definition on the ground that it left unexplained the mechanism by which macro-laws are 'nothing over and above' micro-laws:

how are the macro-laws supposed to come for free? They are not of course logically derivable from the microphysical laws alone, and on Pettit's view they are not reducible to them. And the mere fact that they do not conflict with the microphysical laws does not on its own explain the supervenience picture. So Pettit needs to tell us exactly how the macro and micro 'levels' are connected.<sup>13</sup>

Crane goes on to consider the possibility that the macro- and micro-laws are themselves connected by laws. But since laws connecting, for example, psychology and microphysics will make irreducible mention of properties not appearing in microphysics, Crane thinks that such a move will not be able to block the possibility that the macro–micro connections themselves are non-physical.

<sup>12</sup> P. Pettit, 'A Definition of Physicalism', *Analysis*, 53 (1993), pp. 213–23, at p. 219.

<sup>13</sup> T. Crane, 'Reply to Pettit', *Analysis*, 53 (1993), pp. 224–7, at p. 227.



In these remarks Crane may be seen as raising the possibility that the dependency involved in supervenience might be grounded in an external relation, which relation might be physicalistically unacceptable (so infecting the acceptability of the supervenient property). Pettit responded to Crane with the following analogy, designed to show that facts about supervenient properties can be systematically connected to facts about physicalistically acceptable base properties without being related by laws:

Consider a two-dimensional world in which certain dots constitute shapes of various kinds, and in which no other shape-makers appear, in particular no continuous lines; the theory according to which this is how things are in the world may be called 'dottism'. Under the assumption of dottism, most of us will agree that the shape-configuration of the world in question supervenes on the dot-configuration: duplicate the dot-configuration and you will duplicate the shape-configuration too.<sup>14</sup>

Pettit's analogy proceeds by showing how the relation between the laws involved in physicalist supervenience, like the relation between shape and dot laws, might be an internal relation: that the shape configuration supervenes upon the dot configuration is simply a matter of their natures. Just as there is no external relation holding between the dot laws and the shape laws – set up the dots in a particular configuration and the shape comes along free – there is no external relation holding between laws at different levels generally.

Pettit (p. 144) does not do much to explain how the dot analogy relates to the details of the supervenience in question: 'by my intuitions the shapes–dots dependence robustly illustrates – illustrates independently of further glosses – the sort of dependence that, according to the microphysicalist, the mental and other aspects of the macro-world display in relation to the physical'. Should a further gloss on the dependency relation between properties be desired, however, we can connect Pettit's dot story to a more specific story involving shared causal powers. For Pettit, empirical laws are descriptions of causal relations holding between property instantiations; so to say that macro-laws are nothing over and above micro-laws is plausibly to say that the causal powers of macro-properties are nothing over and above the causal powers of micro-properties. The *reason*, then, why the shapes come free, given the dots, is that the set of causal powers associated with a particular shape configuration stands in an appropriate internal relation to the set of causal powers associated with the dot configuration. Suppose, for example, that a set of dots in a particular configuration has the causal power to remind me of a triangle. The triangular shape supervening on this configuration also has the causal power to remind me of a triangle; but,

<sup>14</sup> P. Pettit, 'Microphysicalism, Dottism and Reduction', *Analysis*, 55 (1995), pp. 141–6, at p. 141.

plausibly, this is not a *different* causal power from that possessed by (the property of being some) dots so arranged; and it seems likely that any causal power possessed by the shape will also be possessed by the dots so arranged. Reference to CCP thus gives substance to Pettit's claim that shape laws are nothing over and above dot laws, and so, by extension, to his claim that macro-laws are nothing over and above micro-laws.

Finally, I consider Sydney Shoemaker's account of realization of higher-level properties by lower-level properties.<sup>15</sup> This account makes explicit how fulfilment of CCP enters into developing a physicalistically acceptable inter-level connection. Its starting-point is Shoemaker's causal theory of properties: 'The claim of the causal theory of properties is that the properties that have causal features non-derivatively [to rule out cases like 'grue'] have them essentially, and are individuated in terms of them'.<sup>16</sup> Shoemaker's causal features (what I have been calling causal powers) may be forward-looking conditional powers (where, roughly, a conditional power of a property is one that, were it to be combined with other conditional powers, none of which is individually sufficient to bestow the causal power in question, could bestow a manifest causal power on the entity possessing the property). Or they may be backward-looking (having to do with how the instantiation of the property was caused).

Shoemaker's suggestion regarding the realization relation is along the lines of Watkins' suggestion (see p. 45 above): what it is for a higher-level property to be realized by a lower-level property is for the set of forward-looking conditional causal powers associated with the higher-level property to be a subset of the set of forward-looking conditional causal powers of the lower-level property. Given this account of realization, it will never be the case that a given higher-level property has a conditional causal power different from any of those of its realizer base property, and so fulfilment of CCP is built into Shoemaker's account. However, it may be the case that the lower-level property has more conditional causal powers than the higher-level property, in which case the subset of conditional causal powers associated with the higher-level property will be a proper subset of the set associated with the realizer property, supporting Shoemaker's non-reductivist claim that a realized property may be distinct from its realizer base property. And it might be the case that the set of conditional causal powers associated with the realized property might be a subset, proper or

<sup>15</sup> S. Shoemaker, 'Realization and Mental Causation', unpublished essay, 1998. A short version of this essay will be forthcoming in the *Proceedings of the 20th World Congress*.

<sup>16</sup> S. Shoemaker, 'Causal and Metaphysical Necessity', *Pacific Philosophical Quarterly*, 79 (1998), pp. 59–77, at p. 65. See also his 'Causality and Properties', in P. van Inwagen (ed.), *Time and Cause* (Dordrecht: Reidel, 1980), pp. 109–35.

improper, of a variety of sets associated with different realizer base properties, supporting his claim that higher-level properties may be multiply realizable.

Shoemaker's (unpublished) account allows for a solution to the problem of causal overdetermination that makes explicit reference to causal powers. As applied to the case of mental properties, the solution is as follows:

When a mental property is instantiated it is one of a set of coincident property instances that form a hierarchy such that instances lower in the hierarchy 'contain' or 'involve' those higher in the hierarchy ... all the instances in the hierarchy are instances of properties that bestow the conditional powers associated with a certain mental property ... the mental instance is sometimes the cause of an effect because it is the conditional powers specifically bestowed by it that contribute to the causing of the effect ... it is a certain set of conditional powers, a subset of all the conditional powers bestowed by the realizing property, that determines what it can cause.

More generally, a higher-level property is not in causal competition with its lower-level realizer base property, since the conditional causal powers of the former are a subset of those of the latter. Shoemaker's account of realization provides a solution to the problem of causal overdetermination because he has built into his account the requirement that the causal powers of the properties involved must stand in an appropriate internal relation, namely, the subset relation. Since the property instances are *coincident*, the conditional causal powers in the subset associated with a higher-level property are not distinct from the conditional causal powers of lower-level properties in the hierarchy, but stand rather in the (internal) relation of part to a single whole.

## V

A supervenience constrained by CCP enables appropriate physicalist solutions to a variety of problems (most importantly, the need for a compatibilist account of the efficacy of properties at different levels in the layered picture). Moreover, it appears likely that any account of supervenience that allows the sets of causal powers of the properties involved to stand in an external relation will be compatible with the supervenient property's being physicalistically unacceptable. Arriving at a physicalistically acceptable supervenience relation has, *pace* Horgan, Levine, Chalmers and others, less to do with an epistemic requirement on explanation than with a metaphysical requirement on causal powers. That fulfilling CCP is what is really at issue in physicalist accounts of supervenience (for cases involving higher- and lower-level properties instantiated in the same individuals) is confirmed by the fact that a wide variety of non-reductive physicalist accounts, both those

that have been discussed here and others, have set up their accounts so that CCP may be met. Moreover, it is clear that reductive physicalist accounts also need to, and do, ensure the fulfilment of CCP, although the focus on identification of properties tends to obscure the fact that they are doing so. Here I hope to have established that these various physicalist accounts are more similar than their presentations make them appear, and that what they share in common must be common to any account of a physicalistically acceptable supervenience relation. It turns out to be misleading to call such a relation ‘supervenience’, much less ‘superdupervenience’; given the constraint that must be put upon it, it would be better termed ‘innervence’.<sup>17</sup>

*Cornell University*

<sup>17</sup> Thanks to Richard Boyd, John Heil, Sydney Shoemaker, Jason Stanley and the participants of John Heil’s 1996 NEH Summer Seminar in the Metaphysics of Mind, for comments, conversation and inspiration. Special thanks to Michael Watkins for making the crucial observation mentioned in fn. 10.