

## PHYSICALISM, NOTHING BUTTERY, AND SUPERVENIENCE

*Giovanna Hendel*

### *Abstract*

I consider the position (which I call ‘the triad’) according to which *physicalism* is a reductive claim which is capturable in terms of the idea (the ‘*nothing buttery*’ idea) that there is nothing but/nothing over and above the physical, an idea which, in its turn, is meant to be capturable in terms of a determinate form of *supervenience*. (Physicalism is then meant to be capturable in terms of the form of supervenience in question.) I argue that there is a tension in the triad. The notion of ‘nothing buttery’ required has features which can’t be captured by the supervenience of the triad. Hence, one cannot have both physicalism as nothing-buttery-reductive and physicalism as supervenience of the kind in question. If one wants to hold onto the idea of physicalism as nothing-buttery-reductive, one must be prepared to identify physicalism with a much stronger claim than one might have originally thought, a claim that can’t be captured by the supervenience of the triad.

### 1. The Triad

There are philosophers who take physicalism to be a reductionist position. This is then meant to be expressed by what I shall call certain ‘nothing buttery’ claims,<sup>1</sup> i.e. by claiming that physicalism is the thesis that there is *nothing but/nothing over and above* the physical, that all particulars and universals are *nothing but/nothing over and above* physical particulars and universals. Finally, considering universals, this ‘nothing buttery’ idea is meant to be capturable by a form of supervenience, to the effect that, within a determinate set of possible worlds, no two worlds differ without differing physically.

We thus have an interplay between three notions: physicalism, ‘nothing buttery’, and supervenience. And these notions are characterized by certain equivalences: physicalism is meant to be

<sup>1</sup> As far as I know, the term ‘nothing buttery’ is to be credited to Jack Smart. (See Smart (1981).)

a reductive claim which is capturable in terms of a nothing buttery claim, capturable, in its turn, by a determinate form of supervenience. (Physicalism is then meant to be capturable in terms of the form of supervenience in question.) In what follows, I'll refer to the three notions as 'the triad', and to the philosophers upholding the kind of position under discussion as 'the triadists'.

The triadists are no strawmen: I think we can regard as triadists philosophers such as Lewis (1983, 1994) and Jackson (1998). That this is true is, of course, best shown by the texts in their whole. There are, however, some specific places which are particularly relevant to support my claim.

For example, on p. 361 of Lewis (1983), the author speaks of Physicalism (or 'Materialism', as he puts it) as a 'reductionist view'. He then characterizes it as

the thesis that physics . . . is a comprehensive theory of the world, complete as well as correct. The world is as physics says it is, *and there's no more to say*. World history written in physical language is all of world history.<sup>2</sup>

The more precise formulation in terms of supervenience comes three pages later:

Among worlds where no natural properties alien to our world are instantiated, no two differ without differing physically; any two such worlds that are exactly alike physically are duplicates.<sup>3</sup>

Similarly, on pp. 412–3 of Lewis (1994), the author characterizes physicalism as a way of expressing a form of 'reductionism about everything' (p. 412). This form of reductionism is first expressed by way of the following claim (C1):

If two possible worlds were exactly isomorphic in their patterns of co-instantiations of fundamental properties and relations, they would thereby be exactly alike *simpliciter*.<sup>4</sup>

Given the physicalist claim that 'all fundamental properties and relations that actually occur are physical' (p. 413), we then get the following claim (C2) as a determinate of C1:

<sup>2</sup> Lewis (1983), p. 361. Italics mine.

<sup>3</sup> Lewis (1983), p. 364. A property is *alien to a world* iff '(1) it is not instantiated by any inhabitant of that world, and (2) it is not analysable as a conjunction of, or as a structural property constructed out of, natural properties all of which are instantiated by inhabitants of that world' (Lewis (1983), *ibid.*).

<sup>4</sup> Lewis (1994), p. 412.

If two worlds were physically isomorphic, and if no fundamental properties or relations alien to actuality occurred in either world, then these worlds would be exactly alike *simpliciter*.<sup>5</sup>

And C2 is, in fact, claimed to express physicalism. Further on, Lewis considers the case of the supervenience of a picture on ‘a million tiny spots – pixels – each of which can be made light or dark’ (p. 413). He says that

The case evokes *reductionist comments* . . . the picture and [its gestalt] properties reduce to the arrangement of light and dark pixels. They are *nothing over and above* the pixels.<sup>6</sup>

And on p. 414 Lewis reiterates that ‘in such a case . . . supervenience is reduction’. He then adds that

the materialist supervenience of mind and all else upon the arrangement of atoms in the void – or whatever replaces atoms in the void – is another such case.<sup>7</sup>

As for Jackson (1998), I’d like to draw the reader’s attention to the claim, in the first paragraph of page 1, that the ‘materialism or physicalism’ he focuses on is a reductive view. He further claims, on p. 9, that

the physicalists’ distinctive doctrine is, as they variously say it, that the world is entirely physical, that it is *nothing but*, or *nothing over and above*, the physical world, and that a full inventory of the instantiated physical properties and relations would be a full inventory *simpliciter*.<sup>8</sup>

The formulation in terms of supervenience then comes on p. 12:

Any world which is a *minimal* physical duplicate of our world is a duplicate *simpliciter* of our world.<sup>9</sup>

In what follows I want to argue that there is a tension in the triad. The notion of ‘being nothing over and above’ that can make of a ‘nothing buttery physicalism’ a reductive position is

<sup>5</sup> Lewis (1994), p. 413.

<sup>6</sup> Lewis (1994), p. 413. Italics mine.

<sup>7</sup> Lewis (1994), p. 414.

<sup>8</sup> Jackson (1998), p. 9. First two italics mine.

<sup>9</sup> Jackson (1998), p. 12. I trust that it is pretty obvious that this claim is equivalent to the following claim (corresponding to my way of formulating the triadists’ supervenience): Within the set of worlds which are minimal physical duplicates of our world, no two differ without differing physically.

to be characterized by features which do *not* characterize a relation that can be captured by the form of supervenience of the triadists. Hence, one cannot have both physicalism as nothing-buttery-reductive and physicalism as what can be adequately expressed by the type of supervenience thesis of the triadist. If one wants to hold onto the idea of physicalism as nothing-buttery-reductive, one must be prepared to identify physicalism with a much stronger claim than one might have originally thought, a claim that can't be captured by the form of supervenience of the triadists.<sup>10</sup>

## 2. From Reduction for Theories to Reduction for Properties

According to the triadist, physicalism, and, hence, the nothing buttery claim which is meant to capture physicalism, is a reductive position. The physicalism of the triadist, however, is – it will be remembered – also a thesis concerning universals such as properties.<sup>11</sup> By contrast, the notion of reduction, at least as originally conceived, concerns theories. It thus seems fair to require from the triadist a notion of reduction which can apply (also) to properties. Moreover, in order for such a notion to be a notion of 'reduction', it is necessary to preserve at least some link with the notion of reduction as originally conceived. (Otherwise one could wonder why, after all, the triadist's notion should be regarded as a notion of *reduction*.) I'll thus begin by considering this latter, and what, if anything, can be built on it.

In Nagel (1961), reduction is defined as 'the explanation of a theory or a set of experimental laws established in one area of inquiry, by a theory usually though not invariably formulated for some other domain.'<sup>12</sup> Explanation is then understood in terms of

<sup>10</sup> Note that this is not to say that physicalism as nothing-buttery-reductive cannot be captured by *any* form of supervenience – nor, for that matter, that one can't be a physicalist, a nothing-butteryist, and a believer in supervenience: the tension that I identify in the triad arises only when one wants one's physicalism to be both nothing-buttery-reductive and what can be adequately expressed by the type of supervenience thesis of the triadist.

<sup>11</sup> See, e.g., Lewis (1983), p. 361: 'Materialism is not just the thesis that there are no things except those recognised by physics. To be sure, materialists don't believe in spirits, or other such nonphysical things. But antimaterialists may not believe in spirits either – their complaint needn't be that physics omits some of the things that there are. They may complain instead that physics overlooks some of the ways there are for physical things to differ; for instance, they may think that physical people could differ in what their experience is like.' Also, see Jackson (1998), p. 6, note 5: 'Our primary concern is with physicalism as a doctrine about the *kind* of world we are in. From this perspective, attribute dualism is no more physicalistically acceptable than is substantial dualism.'

<sup>12</sup> Nagel (1961), p. 338.

derivability: for A to be explainable by B is for A to be derivable from B. And ‘derivability’ is, in turn, understood in terms of Hempel and Oppenheim’s model: A is derivable from B iff  $\square (B \rightarrow A)$ ;  $B ; \therefore A$ .<sup>13</sup> Putting all of that together, a theory or set of laws As is reducible to a theory or set of laws Bs iff, for all A, there exists a law (or cluster of laws) B such that  $\square (B \rightarrow A)$ .

The picture we are offered is, thus, one where what is reduced is ‘assimilated by’/‘incorporated in’ the reducer. (Cf. Nagel (1961), p. 340; and Hempel (1969), p. 190.) In this way, it becomes possible to regard the reducer as capable of explaining not only what is reduced, but also whatever is explainable by what is reduced. Nagel’s model of reduction is a special case of Kemeny and Oppenheim’s.<sup>14</sup>

When one comes to the issue of providing a notion of ‘reduction for properties’, to be somehow related to ‘reduction for theories’, a natural idea is that of understanding ‘reduction for properties’ *in terms of* ‘reduction for theories’: property reduction would be understood as the relation which holds between the properties of two theories when one is reducible to the other. (Cf. Churchland (1986).) In the case of the model/s of theory reduction above, the properties of a theory  $T_1$  would thus be reducible to the properties of a theory  $T_2$  iff  $T_1$  is explainable by/derivable from  $T_2$ ; iff  $T_1$  can be ‘assimilated by’/ ‘incorporated in’  $T_2$ ; iff  $T_2$  can explain all that is explainable by  $T_1$ .

In this way, however, we would have made of property reduction a semantic notion, since theory reduction is typically a semantic notion. (At least in theory reduction as traditionally construed, the relata – i. e. theories/sets of laws – are conceived of as linguistic entities.) The notion of reduction in play in the triad, by contrast, does *not* seem to be semantic.<sup>15</sup> Hence, the

<sup>13</sup> Hempel and Oppenheim (1948).

<sup>14</sup> See Kemeny and Oppenheim (1956). Kemeny and Oppenheim require, for a theory  $T_2$  to be reducible to another theory  $T_1$ , that  $T_1$  ‘can explain all that  $T_2$  can and normally more’. (Kemeny and Oppenheim (1956), p. 13.) They then explicitly claim that Nagel’s model can be taken as a special case of their model. (See Kemeny and Oppenheim (1956), pp. 15 ff.) See also Hooker (1981) and Churchland (1986) for some more recent models where it is possible to find the idea of reduction as the explanation by the reducer of what is reduced and of what is explainable by what is reduced.

<sup>15</sup> Cf. Lewis (1983), p. 358; and Stalnaker (1996) where what is described is, in fact, meant to be the kind of picture subscribed to by Lewis. (Stalnaker (1996), p. 223 and p. 231–232.) Stalnaker distinguishes between reductive and non-reductive notions of supervenience. Commenting on the reductive notion, he says (p. 223–224): ‘... a reductionist thesis was traditionally construed as a semantic thesis: reduction is a relation between theories or languages. One theory is reducible to another if (a) all the terms of one are definable in

notion of property reduction above does not seem suitable for being the notion of reduction in play in the triad.

Another unsuitable candidate is property reduction understood in terms of identity between (simple) properties. For it is obvious that this is not what the triadists have in mind: Why all this talk of ‘nothing buttery’ and ‘supervenience’ if what they are after could be straightforwardly put in terms of identity?<sup>16</sup>

Better, then, turning to a different understanding of property reduction. And as far as I can see, the only option left – given what has just been argued – is that of understanding property reduction itself (and not only *theory* reduction) in terms of derivability: for a family of properties A to be reducible to a family of properties B would be for the As to be derivable from the Bs.

Derivability is, primarily, a logical notion, typically understood as a relation between formulae or (sets of) propositions. What can it mean, then, to speak of ‘derivability’ when the relata are properties? Trying to keep as close as possible to derivability for theories, we can, to begin with, understand derivability for properties thus:

(D) The properties As are derivable from the properties Bs iff, for all A, there exists at least one B in the set of Bs such that, necessarily, if B is instantiated, then A is instantiated.

By understanding property reduction in terms of derivability as in (D), the link with theory reduction (or at least with theory reduction understood as above) is preserved – as it should: both in the case of theory reduction (as above) and in the case of property reduction understood in terms of derivability as in (D), the core idea is that we have conditionals of the form  $\Box (B \rightarrow A)$ , where B stands for a reducer and A stands for a reduced.<sup>17</sup> In order for the

terms of the other, and (b) given the definition, the reduced theory is derivable from the theory to which it is reduced. In contrast, a supervenience relation is supposed to be metaphysical, and not semantic: it is a relation between sets of properties or facts, and not between theories or languages.’ (For a relation to be purely ‘metaphysical’, it then can’t be understood in terms of a relation between theories or languages.) And further on (p. 224): ‘... traditional reductionism makes a metaphysical claim, but also carries with it some metaphysically irrelevant semantic baggage. The concept of supervenience is supposed to be a concept that helps isolate the metaphysical part of a reductionist claim ...’

<sup>16</sup> Note that this is not to say that a triadist who is a physicalist can’t be (also) an identity theorist. The point is just that the notion of property reduction in play in the triad appears to be such as to make of the identity theory only *one* of the possible ways in which one can be a nothing-buttery-physicalist. (Cf. Lewis (1983), p. 361.)

<sup>17</sup> This core idea is, in fact, present in Jackson (1998) when he speaks of the ‘entailment’ of the mental by the physical. (As we are going to see, though, Jackson’s specific way of understanding this core idea is inadequate to preserve the link with theory reduction.) Also, see Lewis (1994), esp. p. 414.

analogy with our notion of theory reduction to be complete, however, it must also be possible to make some sense of the A properties being ‘assimilated by’/‘incorporated in’ the B properties.

This can be done by

(i) conceiving of properties as sets of (actual and possible) particulars,

and by

(ii) conceiving of the derivability of the As from the Bs as requiring, more specifically with respect to (D), that

(D\*) For all instantiation Ax of a property A by an individual x, there exists an instantiation Bx of a property B by x for which it is true that whenever it is instantiated by an individual, A is also instantiated by the relevant individual.

Given (i) and (ii), it will be possible to make sense of the idea of the A properties being ‘assimilated by’/‘incorporated in’ the B properties to the extent that the set of the members of the A properties will be a subset of the set of the members of the B properties.

The idea of ‘assimilation’/‘incorporation’ is not, of course, characteristic of any model of theory reduction. (While it fits well with what is sometimes called ‘conservative/retentive reduction’, it does not seem to fit equally well with notions of ‘eliminative reduction’.) To this extent, one could argue that it is not necessary, to preserve some link with theory reduction, that a notion of property reduction be characterized by the features above (i.e., conception of properties as sets of particulars, and conception of reduction as derivability understood as in (ii)).

However, what remains true also under models of theory reduction which are not associated with the idea of ‘assimilation’/‘incorporation’ of the reduced by the reducer, is, at least, that the reducer must explain all that can be explained by the reduced. And this requirement cannot be satisfied if (D\*) is not. For if (D\*) is not satisfied, room will be left for the possibility of explananda, explainable in terms of As, which are not explainable in terms of Bs. (It will be possible to have, as potential explanantia, instantiations of As which are not instantiations of Bs.) In order to rule out this possibility, to keep the parallel with theory reduction, it is thus necessary to understand the notion of property-reduction-as-derivability in accordance with (ii).

This, together with what has been argued above when considering alternative candidates, means that the best candidate for a triadist notion of property reduction able to preserve a link with theory reduction is to be understood, if not, necessarily, in terms of *both* (i) and (ii), at least in terms of (ii). As we have seen, the nothing buttery relation in play in the triad is meant to be a case of property reduction. (It is meant to capture a reductive form of physicalism which is to concern (also) properties.) Hence, the nothing buttery relation that is needed for the triad is to relate physical and non-physical properties so that for all instantiation  $Mx$  of a non-physical property  $M$  by an individual  $x$ , there exists an instantiation  $Px$  of a physical property  $P$  by  $x$  for which it is true that whenever it is instantiated by an individual,  $M$  is also instantiated by the relevant individual. This – as I am now going to argue – makes the form of supervenience of the triadist unsuitable to capture the nothing buttery relation that is needed for the triad. (This relation will, henceforth, be called ‘nothing-buttery-reduction’ – or ‘NBR’, for short.)

### 3. Nothing-Buttery-Reduction and Supervenience

The core idea of supervenience is that sameness in certain respects (the *subvenient* respects) yields sameness in certain other respects (the *supervenient* respects). A form of supervenience can then be *global* or *local*: it is global if what is compared, when speaking of ‘sameness’, are entire possible worlds; it is local if what is compared, when speaking of ‘sameness’, are single individuals.

Among local varieties of supervenience features the following form:<sup>18</sup>

$$(SS) \quad \Box (\forall x) (\forall F \in A) \{Fx \rightarrow (\exists G \in B) [Gx \ \& \ \Box (\forall y) (Gy \rightarrow Fy)]\}$$

When the two modal operators are understood as indicating that what falls within their scopes is to be taken as an absolute necessity (i.e. as a truth holding in all possible worlds), (SS) is equivalent to (D\*). Given what has been argued in the previous section, we can thus say that there is at least one form of local supervenience that is adequate to express NBR. In particular, then, (SS) allows us – as wished – to tie property reduction to theory reduction insofar as, when (SS) holds, we’ll have, for every instantiation  $Fx$  of an A property, an instantiation  $Gx$  of a B property and a relation

<sup>18</sup> See Kim’s ‘strong supervenience’ as formulated in, e.g., Kim (1993), p. 65.

expressed by the universal conditional  $\Box (\forall y)(Gy \rightarrow Fy)$  which yields  $Fx$  as a result of  $Gx$ ; it will thus be possible, for every instantiation  $Fx$  of an A property, to mimic theory reduction by claims of the following form (for properties G and F as in (SS)):

$$(C) \Box (\forall x)(Gx \rightarrow Fx); Gx; \therefore Fx$$

Considering global forms of supervenience, it is not, by contrast, clear that the truth of claims of type (C) is guaranteed. For to say that what is compared – when speaking of ‘sameness’ in our formulation of the core idea of supervenience – are, in the case of global forms, entire possible worlds, is to say that the supervenient properties of a determinate individual are not, necessarily, going to be correlated with subvenient properties of this same individual.

The form of supervenience of the triadists (henceforth, ‘TS’) is global. (As it will be remembered, the idea was that, within a determinate set of possible worlds, no two worlds differ without differing physically.) To the extent that what has just been pointed out can be taken to be a problem for capturing NBR in terms of a global form of supervenience, it is thus going to be a problem for capturing NBR in terms of TS. (The idea is that if global forms of supervenience can’t guarantee the truth of claims of type (C), then they can’t guarantee (D\*) either, for – as we have just seen – (D\*) entails the truth of claims of type (C).)

As Kim and others have shown,<sup>19</sup> however, global forms of supervenience are equivalent to local forms where the subvenient and the supervenient sets are closed under identity and quantification. (The closure of the subvenient and supervenient sets under identity and quantification allows for the construction of relational properties such as  $Px = P'x \ \&\ \ (\exists y) (x \neq y \ \&\ P'y)$  which, in formulations of local forms of supervenience such as (SS), can play the role of what was referred to by ‘G’ in our formulation. That makes it possible for a formulation of supervenience such as (SS) to cover cases where the instantiation of a supervenient property F by an individual x covaries with instantiations of B properties by individuals other than x.) So, as long as the subvenient and the supervenient sets are closed under identity and quantification, claims of type (C) can, after all, be entailed by forms of global supervenience.

Still, the closure of the subvenient and supervenient sets under identity and quantification might, actually, be unacceptable to

<sup>19</sup> See Kim (1993), p. 169; Klagge (1995); Stalnaker (1996).

many. (Think of what such a closure would mean. When, for example, the subvenient set is the set of the physical properties, such a closure would require one to count as physical not only a property such as that of having a determinate physical property  $P$  and a world-mate who instantiates another physical property  $P'$ , but also a property such as that of merely having a world-mate who instantiates a physical property  $P$  – a property which could be had by a Cartesian soul.) And if one rejects the closure of the subvenient and supervenient sets under identity and quantification, the global character of TS is, indeed, going to be a problem for the possibility of its capturing NBR. (If one rejects the closure of the subvenient and supervenient sets under identity and quantification, it would no longer be possible, for global forms of supervenience, to guarantee the truth of claims of type (C), and, hence, as we have seen, of (D\*.) All that I wanted to concede is thus that, in the case of worries arising from the global character of TS, there *might* be a way out. In the case of worries arising from the fact that TS is characterized by a restricted modality, by contrast, this does not – as it is going to be clear from what follows – seem to be the case.

The holding of conditionals associated with TS is restricted to a determinate set of possible worlds.<sup>20</sup> This can be put by saying that TS is characterized by a restricted modality. The motivation behind this feature is that of wanting TS to be consistent with possible worlds where – as Jackson puts it for the psychophysical case – there is, in addition to mental life sustained in material stuff, a lot of mental life sustained in non-material stuff. In other words – and focusing, once again, on the psychophysical case – TS is to be consistent with possibilities such as that of having, in addition to instantiations of mental properties by physical individuals, also instantiations of mental properties by individuals who are *not* physical. (This kind of possibility will be called ‘the possibility of dualism’, or ‘PD’ for short.) Now the triadist wants his physicalism to allow for PD. Given this picture of physicalism, it is thus necessary for TS to be consistent with PD, if TS is to be suitable to capture physicalism.

<sup>20</sup> The conditionals in question are conditionals between instantiations of complex relational properties, of the form  $Px = P'x \ \& \ (\exists y) (x \neq y \ \& \ P'y)$ , corresponding to the entire distributions of physical and non-physical properties of our world. (Cf. Jackson (1998), p. 25). The determinate set of possible worlds in question is, according to the version of TS considered, either the set of the minimal physical duplicates of our world, or the set of possible worlds where – as Lewis puts it – there are no properties alien to our world.

But PD is not consistent with NBR. For the possibility of having instantiations of properties to be reduced by individuals who do not possess any of the properties which are to count as reducer is inconsistent with (D\*), and thus, with a claim, such as NBR, which, as we have seen in the previous section, is to be understood in terms of (D\*). So, TS's consistency with PD – and that feature of TS, i.e. its restricted modality, which makes this consistency possible – makes TS unsuitable to capture NBR.

Cases of NBR can't, then, be identified with cases of TS. What can be captured by TS can't, after all, be the same as what can be captured by a nothing buttery claim of the kind considered. And if physicalism is to be a case of NBR, then physicalism can't be captured by TS. Reduction can't be bought that cheap.

*Department of Philosophy*  
*University College London*  
*Gower Street*  
*London WC1E 6BT*  
*UK*  
*G.Hendel@ucl.ac.uk*

## References

- Churchland, P. (1986) *Neurophilosophy* (Cambridge Massachusetts: MIT Press).
- Hempel, C. (1969). 'Reduction: Ontological and Linguistic Facets'. In S. Morgenbesser, P. Suppes and M. White (eds.), *Philosophy, Science, and Method*, pp. 179–199 (New York : St. Martin's Press).
- Hempel, C. and P. Oppenheim (1948). 'Studies in the Logic of Explanation', *Philosophy of Science* 15: 135–175.
- Hooker, C. (1981). 'Towards a General Theory of Reduction', *Dialogue* 20: 38–59; 201–236; 496–529.
- Jackson, F. (1998). *From Metaphysics to Ethics* (Oxford: Clarendon Press).
- Kemeny, J. G. and P. Oppenheim (1956). 'On Reduction', *Philosophical Studies* 7: 6–19.
- Kim, J. (1993). *Supervenience and Mind* (Cambridge: Cambridge University Press).
- Kluge, J. C. (1995). 'Supervenience: Model Theory or Metaphysics?' In E. E. Savellos and U. D. Yalçin (eds.), *Supervenience: New Essays* (Cambridge: Cambridge University Press).
- Lewis, D. (1983). 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* 61: 343–377.
- Lewis, D. (1994). 'Lewis, David: Reduction of Mind'. In S. Guttenplan (ed.) *A Companion to the Philosophy of Mind* (Oxford: Blackwell).
- Nagel, E. (1961). *The Structure of Science* (London: Routledge & Kegan Paul).
- Smart, J. J. C. (1981). 'Physicalism and Emergence', *Neuroscience* 6: 109–113.
- Stalnaker, R. (1996). 'Varieties of Supervenience', *Philosophical Perspectives* 10: 221–241.