FRANZ VON KUTSCHERA

SUPERVENIENCE AND REDUCTIONISM

Materialism is a reductionistic position. It claims that in the last analysis all things and events are physical objects and events, that the language of physics is sufficient to express all empirical facts, or that "the world is as physics says it is, and there's no more to say".[1] For about 20 years, however, non-reductive versions of Materialism have now dominated the scene, i.e. non-reductive reductionisms. This sounds puzzling, but "reduction" has to be understood in a different sense in the two components. "Non-reductive" means that the theory does not imply definability. The sense in which it is still reductionist is less clear. In one version non-reductive Materialism says that there is a relation between non-physical and physical properties – supervenience – that does not imply reduction as definability, but nevertheless supports a materialist conception of reality. In another version it says that all events are identical with physical events.[2] This claim, however, makes sense only with respect to a theory of events, and since I am not satisfied with present theories I shall be concerned only with the first version of non-reductive Materialism in this paper. The central question will be: Is supervenience really a sufficient basis for Materialism? Since different concepts of supervenience have been discussed in the literature and their properties and relations are not always correctly understood, we first have to tackle conceptual matters.

1. CONCEPTS OF SUPERVENIENCE

Let F and G be families of properties defined on a common domain D of objects, and let $f, f', \ldots$ be elements of F, $g, g', \ldots$ those of G and $x, y, \ldots$ those of D. Then a first concept of supervenience is defined by:

D1: F is *supervenient* on G iff $\Lambda xy(\Lambda g(gx \equiv gy) \supset \Lambda f(fx \equiv fy))$,

i.e. iff there is no difference in F-properties without a difference in G-properties. As stated this is a concept of *extensional* supervenience which is of no interest for reductionistic purposes. If we put an operator

Erkenntnis **36**: 333–343, 1992.

for necessity before the definiens we obtain a notion of *nomological supervenience*, if the operator expresses truth in all worlds in which the same natural laws hold as in our world, and a notion of *analytical supervenience*, if it expresses truth in all possible worlds. Since supervenience is usually taken as logically contingent, nomological or, as it is mostly called: *weak supervenience* (WS), is the most interesting case for our discussion.

Supervenience is often equated with inferentiality:

D2: F is *inferential* with respect to G iff $\Lambda xf(fx \supset Vg(gx \wedge \Lambda y(gy \supset fy)))$,

i.e. iff an object has property f only in virtue of having property g, which is a sufficient general criterion for being f. Inferentiality implies supervenience, but we can only say: If F is supervenient on G then F is inferential with respect to G^+, where G^+ is the smallest complete Boolean algebra containing G, i.e. G^+ is the smallest set of properties containing G closed with respect to negation and finite as well as infinite conjunctions. Supervenience, therefore, is equivalent with inferentiality only if $G = G^+$.[3] In this case WS is equivalent to nomological inferentiality, defined by putting an operator N for nomological necessity before the definiens in D2.

Given $G = G^+$ WS is also equivalent with $\Lambda f\mathrm{N}Vg\Lambda x(fx \equiv gx)$. This is essentially weaker than $\Lambda fVg\mathrm{N}\Lambda x(fx \equiv gx)$. And that is what – for $G = G^+$ – *strong supervenience* (SS) amounts to:[4]

D3: F is *strongly supervenient* on G iff $\mathrm{N}\Lambda xf(fx \supset Vg(gx\mathrm{N}\Lambda y(gy \supset fy)))$.

Here, against nomological inferentiality, there is a cross-world connection between the properties g and f, so that being g is a criterion for being f in the strong sense that, if an object y were a g, it would also be an f. SS, then, implies WS, but the converse does not hold.

A third concept of supervenience is that of global supervenience (GS).[5] Let w, w' be nomologically possible worlds as seen from the standpoint of our world and $f_w x$ be the proposition that property f applies to x in world w. Then GS is defined by:

D4: F is *globally supervenient* on G iff $\Lambda ww'(\Lambda gx(g_w x \equiv g_{w'} x) \supset \Lambda fx(f_w x \equiv f_{w'} x))$.

That is: Worlds that are alike with respect to the distribution of G-

properties are also alike with respect to the distribution of F-properties.

D4 can be extended to a notion of supervenience of propositions (or states of affairs): If W is the set of all possible worlds, not just the nomologically possible ones, and X and Y are sets of propositions p and q, respectively, i.e. sets of subsets of W, we say: X is *P-supervenient* on Y iff $\Lambda ww'(\Lambda p(w \in p \equiv w' \in p) \supset \Lambda q(w \in q \equiv w' \in q))$.[6] If X is the smallest complete Boolean algebra containing the propositions that fx for all $f \in F$ and $x \in D$, and likewise for Y and G, then GS of F on G is equivalent with P-supervenience of X on Y for $G = G^+$ and $F = F^+$.

SS implies GS, but the converse does not hold. GS does not imply WS, and WS not GS.

WS, SS and GS are the three most important concepts of supervenience in the current discussion, but they are not the only possible ones, of course. A definition of supervenience concepts for *attributes* (relations besides properties) instead of just properties is not essentially more general. For if F and G are families of attributes defined on D supervenience of F on G is supervenience of the set of properties definable from F by substitution on the set of properties so definable from G. But since we have:

WS: $\Lambda wxy(\Lambda g(g_w x \equiv g_w y) \supset \Lambda f(f_w x \equiv f_w y))$

GS: $\Lambda ww'(\Lambda gx(g_w x \equiv g_{w'} x) \supset \Lambda fx(f_w x \equiv f_{w'} x))$

SS: $\Lambda ww'xy(\Lambda g(g_w x \equiv g_{w'} y) \supset \Lambda f(f_w x \equiv f_{w'} y))$ (for $G = G^+$),[7]

it is natural to add

LS: $\Lambda ww'x(\Lambda g(g_w x \equiv g_{w'} x) \supset \Lambda f(f_w x \equiv f_{w'} x))$

to this list. Let us call this *local supervenience*.[8] LS says: If an object x has the same G-properties in two nomologically possible worlds, it also has the same F-properties in these worlds, i.e. no change in F-properties without a change in G-properties. Since LS, for $G = G^+$, is equivalent with $\Lambda fx\mathrm{V}g\mathrm{N}(fx \equiv gx)$ it also states cross-world correspondences like SS, though not universal ones. Clearly LS implies GS, and is implied by SS. But it neither implies WS, nor is it implied by WS.

The relations between these different supervenience concepts depend on the families F and G. If, for instance, G contains for all $g \in G$ and $x \in D$ the property $g^{\wedge}(y) \equiv y = y \wedge g(x)$, i.e. the property applying to any y iff g applies to x, GS of F on G implies LS of F on G, and GS and WS imply SS.[9]

As usual I have defined all supervenience concepts as relations between properties (or propositions). They can, of course, also be formulated for predicates (or sentences), but it is more convenient to abstract from the structure of languages. Definability and reducibility, on the other hand, are essentially linguistic relations. Let F and G now be disjunct sets of one-place predicate constants of a language L. Then F is *definable* from G in L iff for all $f \in F$ there is a predicate $A[x]$ in L, containing only predicate constants from G, such that $\Lambda x(fx \equiv A[x])$ is true in L. F is analytically (nomologically) definable from G, if this equivalence is an analytical (nomological) truth. Clearly nomological definability of F from G implies SS of the set F' of properties expressed hy the predicates of F on G', the set of properties expressed by the predicates from G. The converse, however, is not true, even though we have $\Lambda f \mathrm{V} g \mathrm{N} \Lambda x(fx \equiv gx)$ for $G' = G'^{*}$. For an $f \in F'$ the corresponding property g from G' is constructed as follows: Let g'_x be the (infinite) conjunction of all g' such that $g'_w x$. Then g is the (infinite) disjunction of all the g'_x such that $f_w x$. Here g is defined by f, so that f cannot in turn be defined by g, and, furthermore, in L we generally don't have the means to build infinite conjunctions of arbitrary sets of predicates, i.e. sets also for which there is no defining predicate in L. So Kim's suggestions notwithstanding,[10] even SS, the strongest form of supervenience here considered, does not imply definability and does not even come "close to it" or shows that it is "possible in principle".

Reducibility is mostly understood as implying nomological definability. A wider notion of reducibility does not presuppose an identity of the predicate domains. If S_1 and S_2 are two languages of the same logical type (languages of first order predicate logic, e.g.) with the interpretations $\mathfrak{I}_1 = \langle W, R, D_1, V_1 \rangle$ and $\mathfrak{I}_2 = \langle W, R, D_2, V_2 \rangle$ (W is the set of possible worlds, R a relation of nomological accessibility on W, D_1 and D_2 are the universes of discourse and V_1, V_2 interpretation functions), then in w_0 S_1 is nomologically reducible to S_2 iff adding the constants of S_1 to S_2 there are explicit definitions D for them in S_2, such that for the extension $\mathfrak{I}_{2D}$ of $\mathfrak{I}_2$ satisfying D we have $V_{2Dw}(A) = V_{1w}(A)$ for all w such that $w_0 R w$. Reducibility in this sense implies P-supervenience of the propositions expressible in S_1 on those expressible in S_2, but again the converse does not hold.

Supervenience then, even in the strongest form here considered does not imply reducibility. The assumption of a supervenience of the non-physical on the physical, therefore is indeed non-reductive. But is it a

materialist conception? As a first step let us consider the plausibility of supervenience claims.

2. THE PLAUSIBILITY OF UNIVERSAL SUPERVENIENCE ASSUMPTIONS

Hitherto I have considered supervenience theses as statements about nomological necessities. As such they would be empirical. Mostly, however, they are seen as metaphysical assumptions, and that, in the case of Materialism as a claim of the supervenience of all properties defined for empirical objects on physical ones, seems much more appropriate. In fact I think that some of these claims have a high apriori plausibility, but that this is a good argument against classifying them as genuine materialist positions.

Let F now be the set of all properties, while G, as before, is the set of all physical properties. Then WS of F on G says:

(1) $\quad N\Lambda xy(\Lambda g(gx \equiv gy) \supset \Lambda f(fx \equiv fy)).$

According to Leibniz's principle

(L1) $\quad \Box\ \Lambda xy(\Lambda f(fx \equiv fy) \supset x = y)$

($\Box$ stands for analytic necessity) this implies

(P1) $\quad N\Lambda xy(\Lambda g(gx \equiv gy) \supset x = y).$

(P1) says that for all empirical objects there are sufficient physical criteria of identity. This even a dualist will not doubt, if he does not assume a Cartesian Dualism of substances and envisages no unembodied spirits. For him it are persons, who have mental properties and for their identity being in the same place at the same time is sufficient. Since (1) follows from (P1) by the principle of substitutivity of identicals, the inverse of (L1), he will also accept a WS of all properties, especially psychological properties, on physical ones.

It has often been said that in cases where (P1) holds a WS on G is trivial and that an interesting ontological position would result only if (P1) is not true. But a universal WS of all properties on physical ones does imply (P1), and it is only plausible in view of this principle.

Now in (P1) the operator N can also be understood as a metaphysical necessity since it expresses our conception of empirical objects, which is independent of our assumption of specific natural laws: Empirical objects always have physical properties and can be identified by them. But then (1) is not just a nomological truth either, but has the same status as (P1).

The case of GS is not quite as simple. It says

(2) $\Lambda ww'(\Lambda gx(g_w x \equiv g_{w'} x) \supset \Lambda fx(f_w x \equiv f_{w'} x))$.

Here we can use the principle:

(L2) $\Lambda ww'(\Lambda fx(f_w x \equiv f_{w'} x) \supset w = w')$.

It says that in different worlds at least one object must have at least one different property, which is highly plausible. For in what sense should the worlds be different otherwise? With (L2) we obtain from (2):

(P2) $\Lambda ww'(\Lambda gx(g_w x \equiv g_{w'} x) \supset w = w')$.

That is: Worlds in which all objects have exactly the same physical properties are identical. This again might be accepted also by a dualist of non-Cartesian profession. First, (P2) does not imply that changes in the mental states of a person are always connected with neurological changes in his brains, but only that these psychological changes are accompanied by some physical changes somewhere in the universe, and this is an extremely weak assumption. Second, the dualist may acknowledge that for worlds there are sufficient physical criteria of identity as there are for persons. Since (2) is again a logical consequence of (P2), he will then also accept (2). However, he might also argue that the assumption that changes in our beliefs about physical objects are always accompanied by changes in them implies implausible restrictions as to what we can know about them.

Things are still more problematic in the case of SS. It says

(3) $\Lambda ww'xy(\Lambda g(g_w x \equiv g_{w'} y) \supset \Lambda f(f_w x \equiv f_{w'} y))$.

Here we could postulate

(L3) $\Lambda ww'xy(\Lambda f(f_w x \equiv f_{w'} y) \supset w = w')$.

That is: Two objects never have exactly the same properties in different worlds. This, however, is only plausible if we consider not just intrinsical

but also extrinsical, i.e. relational properties belonging to an object only in virtue of its relations to other objects with certain properties, like being taller than John Jones. The properties $f^{\vee}$, that apply to every object y just in case f applies to x are such relational properties. If for all $x \in D$ and $f \in F$ $f^{\vee}$ is in F, $\Lambda f(f_w x \equiv f_{w'} y)$ implies $\Lambda f x(f_w x \equiv f_{w'} x)$, so that (L3) is a consequence of (L2). From (L3) and (3) we obtain

(P3) $\quad \Lambda w w' x y(\Lambda g(g_w x \equiv g_{w'} y) \supset w = w')$,

i.e. worlds in which two objects have exactly the same physical properties are identical. Here (3) is not a consequence of (P3), however, since the inverse of (L3) would imply that in all worlds there is just one object. But (P3) entails that SS is a consequence of WS.

Now (L3) and (P3) are highly problematic. But a supervenience that would hold only for intrinsic properties would certainly be too weak for Materialism. There are non-physical properties like being money that certainly do not correspond to intrinsical physical ones.[11] If we admit relational properties however, it will be hard to exclude properties like $f^{\vee}$.[12] But suppose this problem is solved. Then SS would be the best candidate for a distinctive materialist conception of reality. To be a successful candidate, however, it should also imply a dependence of all psychological states on physical ones.

3. SUPERVENIENCE AND DEPENDENCE

Supervenience is often seen as a relation of dependence or determination. Now a dependence of all psychological on physical states or events is something a dualist has to deny, if he doesn't want to water down his position too much. So we have to check whether supervenience indeed implies dependence.

First, however, two remarks. Dependence is an asymmetric relation. But F's supervenience on G does not preclude G's supervenience on F – every property is supervenient on itself, e.g. Therefore we consider only cases, in which an inverse supervenience is excluded. F, then, cannot now be a set including G, as in the last section. We will now take F to be the set of mental properties. Dependence, furthermore, in the context of metaphysical claims has to be understood as a relation between a fact and its *ratio essendi*, i.e. its cause. The materialist does not deny that psychological states may be symptoms and therefore

rationes cognoscendi of brain states, but he maintains that brain states are the causes of mental states – if he doesn't, like D. Davidson, conceive of them as identical.

Now the supervenience concepts here discussed contain no element that justifies an understanding as dependence relations. An extensional supervenience of F on G means just that the most specific partition of D-objects by G-properties is a (proper or improper) subclassification of every partition by F-properties. Truth, to use an example of Davidson, is defined for sentences. That is, (eternal) sentences with different truth values must be syntactically different. Here we even have an analytical supervenience, but nobody would say that the truth of a sentence depended on its syntactical properties, that they of themselves made a sentence true or false. What makes a sentence true is rather the meaning we attach to it together with the facts of the world. J. Kim has also pointed out in (1990) that, since WS correlates a property f to a property g only in one world, we cannot say that if x were g it would be f. As this is a minimum requirement for causal dependence, he advocates SS instead of WS as a materialist thesis. Let us then take this strongest form of supervenience. The argument that it implies a dependence of the distribution of F-properties on that of G-properties is just that it fulfills the minimum requirement: For every f there is a g such that $N\Lambda x(fx \equiv gx)$, if $G = G^+$, and therefore we can say: If x would (not) have had the property g it would (not) have been an f. GS for Kim is too weak as a materialist position, since it does not imply a connection between the physical and the non-physical states of one and the same object. The minimum requirement is not sufficient for causal dependence, however. $N\Lambda x(fx \equiv gx)$ does not say that being f *depends* on being g; it may well be the other way round. Having g can also be a symptom of having f, one that occurs only in cases of f. And $N\Lambda x(fx \equiv gx)$ may also obtain if fx and gx have a common cause.[13] A cause, furthermore, precedes its effect, while supervenience relations correlate coinciding events. Finally, by the nomological conception of causality, ga is only a cause of fa, if there is a law under which both events may be subsumed. But in view of the construction of g $N\Lambda x(gx \supset fx)$ will not in general be a lawlike sentence; there may be no predicate expressing g. Therefore SS neither implies that mental phenomena can be physically explained – even if we assume that some day all physical phenomena are amenable to explanation. What is

needed for a derivation of a sentence "fx" about a mental property of some person x from sentences about physical laws and conditions is not just that for f there is a property g such that $N\Lambda x(fx \equiv gx)$, but that there is a predicate for this g. If there is such a predicate we could define f by it. But SS does not imply definability.[14]

Dualism maintains: Neither can the mental be reduced to the physical, nor the physical to the mental. According to the sense in which "reduced" is understood different dualistic theses can be distinguished. If reduction is taken in the sense of analytical or nomological definability or nomological reducibility in the sense explained at the end of Section 1, forms of Dualism emerge that would be accepted by non-reductive Materialism. For non-Cartesian dualists conflict will also not arise if reducibility is taken as supervenience. To be sure, since a supervenience of the physical on the mental is not at all plausible, the dualist can no longer keep up his symmetrical attitude towards Materialism and Idealism, but this for him is not really important.[15] What he has to defend against materialism is that mental phenomena – or at least some of them – are phenomena in their own right. In one sense this is true already if logical physicalism is wrong, as everyone admits nowadays. Mental phenomena are phenomena described in psychological terms and if these descriptions cannot be translated into the language of physics, they are beyond what physics can ascertain or explain. In another sense, however, mental phenomena are phenomena in their own right only if they are not wholly dependent on physical events. But since supervenience relations do not imply such a dependence, they pose no problem for a dualist. He may accept them and still argue that materialism is false. And that means that supervenience is inadequate for a formation of materialism. To borrow an expression from Kim in (1990): Supervenience is only a relation of *covariance*, not of dependence.

An argument for a one-sided causal dependence of mental on physical events has to proceed from other premisses, therefore. Physicalists – in contradistinction to physicists – generally believe that the physical is causally closed. This assumption, however, is not an argument for, but against a causal dependence of the mental on the physical. It precludes not only non-physical *causes* of physical events but also non-physical *effects* of physical events. Physical causation implies a transfer of energy, and a closed system preserves its energy. The consequence

would then be Parallellism: The nomological correlation between mental and physical events would be correspondences between causally disjunct systems as in Leibniz's example of the two clocks that, once set aright, independently always show the same time.

NOTES

[1] D. Lewis in (1983), p. 361.

[2] Cf. Davidson (1970). Davidson there treats both versions as equivalent, but that presupposes an appropriate theory of events.

[3] Cf. Kim (1978).

[4] Cf. Kim (1984). Nomological necessity is a S5-necessity. The Barcan-formula and its inverse hold since we assume a common domain of possible objects for all worlds.

[5] Cf. Hellman and Thompson (1975).

[6] Cf. Haugeland (1982).

[7] For the equivalence with D3 cf. Kim (1990).

[8] Wether LS deserves its name depends on the properties in G and F. If only intrinsic properties are admitted, LS is indeed local.

[9] So the inclusion of extrinsic properties in G is much more problematic than an extension of G to G^+. For the latter problem cf. Post (1984) and Teller (1984).

[10] Cf. Kim (1978) and (1990).

[11] The example is J. Fodor's. See also Teller (1984).

[12] For another example of a concept with the same consequences cf. T. Horgan (1982).

[13] Cf. Grimes (1988).

[14] While in (1990) Kim doesn't maintain anymore that SS of F on G implies a dependence of the F-phenomena on the G-phenomena, he did so in his earlier papers. There he also argued for a dependence of the causal connections among the former upon those of the latter. In (1984) he introduces a concept of *supervenient causation*: fc causes $f'c$ superveniently if there are supervenience bases g and g' for f and f' (i.e. if we have $N\Lambda x(gx \supset fx)$ and $N\Lambda x(g'x \supset f'x)$) applying to c and being causally connected by a law $\Lambda x(gx \supset g'x)$. In Kim's examples for supervenient causation (the sequence of images, macrophysical causation) fc is always caused by gc, and $f'c$ by $g'c$. In this case talk of a merely epiphenomenal causation makes sense, for then fc is not an independent cause of $f'c$ and it is not necessary to assume a genuine causal relation between fc and $f'c$ to account for their connection. As we have emphasized, however, the relation between a supervenience base and the supervenient property is no causal relation. Hence the following situation is also a case of supervenient causation according to the definition: c is person who has contracted disease f, which causes a reaction f' in the immunological system of c. g is an infallible symptom for f which causes a specific pain g' that only occurs when this symptom is present. Then g' is a nomologically sufficient condition for f' – no pain g' without symptom g, no symptom g without disease f, no disease f without the effect f'. But in this case there is no intuitive justification for saying that the causal relation between disease f and reaction f' is only epiphenomenal with respect to the causal relation between symptom g and pain g'.

[15] He might, however, argue that there is, after all, also a kind of supervenience of the physical on the mental. There is no sense in postulating physical differences that cannot,

in principle be ascertained by observation. Physical differences are not always realized by someone, i.e. there may be physically different worlds with equal distributions of mental attitudes, but for every physical state of affairs there has to be a mental state – its observation by somebody – which implies it nomologically. The supervenience concept which applies here can be obtained by weakening P-supervenience: P-supervenience of X on Y (for $Y = Y^+$) obtains iff $\Lambda p(p \in X \supset Vq(q \in Y \wedge p =_N q)$, where $p =_N q$ is $p \subset_N q \wedge q \subset_N p$ and $p \subset_N q := \Lambda w(w \in p \supset w \in q)$. Weak P-supervenience obtains in case of $\Lambda p(p \in X \supset \mathrm{V}q(q \in Y \wedge q \subset_N p))$, or equivalently $\Lambda w \mathrm{V} w' \Lambda w''(\Lambda q(w'' \in q \equiv w' \in q) \supset \Lambda p(w'' \in p \equiv w \in p)$. In physics we have a similar situation: Mostly only the supervenience of macrophysical attributes on microphysical ones is emphasized, but as P. Suppes has pointed out in (1985) we can also see it the other way round.

REFERENCES

Davidson, D.: 1989, 'Mental Events', in L. Foster and J. W. Swanson (eds.), *Experience and Theory*, Amherst 1970; Reprinted in Davidson: *Essays on Action and Events*, Oxford, pp. 207–27.

Grimes, T. R.: 1988, 'The Myth of Supervenience', *Pacific Philosophical Quarterly* **69**, 152–60.

Haugeland, J.: 1982, 'Weak Supervenience', *American Philosophical Quarterly* **19**, 93–103.

Hellman, G. and F. Thompson: 1975, 'Physicalism: Ontology, Determination, and Reduction', *Journal of Philosophy* **72**, 551–64.

Horgan, T.: 1982, 'Supervenience and Microphysics', *Pacific Philosophical Quarterly* **63**, 29–43.

Horgan, T. (ed.): 1984, *Supervenience*, Memphis/Tenn.

Kim, J.: 1978, 'Supervenience and Nomological Incommensurables', *American Philosophical Quarterly* **15**, 149–56.

Kim, J.: 1984, 'Concepts of Supervenience', *Philosophy and Phenomenological Research* **45**, 153–76.

Kim, J: 1990, 'Supervenience as a Philosophical Concept', mimeographed, University of Bielefeld.

Lewis, D.: 1983, 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* **61**, 343–77.

Post, J.: 1984, 'Comment on Teller', in T. Horgan, pp. 163–67.

Suppes, P.: 1985, 'Davidson's Views on Psychology as a Science', in B. Vermazen and M. B. Hintikka (eds.), *Essays on Davidson*, Oxford pp. 183–94.

Teller, P.: 1984, 'Comments on Kim's Paper', in T. Horgan, pp. 57–62.

Manuscript submitted December 19, 1990
Final version received September 10, 1991

Institut für Philosophie
Universität Regensburg
Universitätsstr. 31
8400 Regensburg