4

Concepts of supervenience

For Herbert Heidelberger

I. INTRODUCTION

We think of the world around us not as a mere assemblage of unrelated objects, events, and facts, but as constituting a system, something that shows *structure,* and whose constituents are connected with one another in significant ways. This view of the world seems fundamental to our scheme of things; it is reflected in the commonplace assumption that things that happen in one place can make a difference to things that happen in another in a way that enables us to make sense of one thing in terms of another, infer information about one thing from information about another, or affect one thing by affecting another. Central to this idea of interconnectedness of things is a notion of *dependence* (or, its converse, *determination*): things are connected with one another in that whether something exists, or what properties it has, is *dependent on,* or *determined by,* what other things exist and what kinds of things they are. It is in virtue of these dependency or determinative relationships that the world can be made intelligible; and by exploiting them we are able to intervene in the course of events and alter it to suit our wishes. Activities like explanation, prediction, and control would make little sense for a world devoid of such connections. The idea that "real connections" exist and the idea that the world is intelligible and controllable are arguably equivalent ideas.

Causation is a preeminent example of what I am calling determinative or dependency relations; apart from those that are logically based, such as entailment, it is the only explicitly recognized and widely discussed relation of this kind. Causes determine their effects, and effects are dependent, for their existence and properties, on their causes. It is not for noth-

An earlier version of this paper was presented at the Herbert Heidelberger Memorial Conference at University of Massachusetts and Smith College in April, 1983. I dedicate this paper to Herb's memory.

ing that Hume called causation "the cement of the universe";[1] causation is the cosmic glue that binds discrete objects and events together, making them mutually significant – even in Hume's atomistic world – and thereby helping to provide a necessary basis for the prediction and control of natural phenomena. It is congenial to the broadly realist view of the world that most of us accept to think of the network of causal relations in the world as underlying, and supporting, the network of explanatory and other epistemic relations represented in our knowledge of it.

The part–whole relation is also important; however, its importance seems to derive largely from the belief that many crucial aspects of a whole including its existence and nature are dependent on those of its parts. That is, mereological relations are significant because mereological determination, or "mereological supervenience,"[2] is, or is thought to be, a pervasive fact.

There has lately been an increasing interest in the concept of *supervenience,* especially for its possible applications to the mind–body problem, microreduction, and physicalism. It is useful to think of supervenience as belonging in that class of relations, including causation, that have philosophical importance because they represent ways in which objects, properties, facts, events, and the like enter into dependency relationships with one another, creating a system of interconnections that give structure to the world and our experience of it. Modes of dependency or determination may differ from one another in various ways; if supervenience is thought of as such a mode, questions arise as to exactly how it differs from others, whether it is a single homogeneous relation or represents in reality two or more distinguishable relationships, and whether supervenient determination presents a philosophically significant alternative to other determinative relations.

The idea of supervenience seems to have originated in moral philosophy. In the following well-known passage G. E. Moore describes a certain dependency relationship between moral and nonmoral properties that has later come to be called "supervenience":

> . . . if a given thing possesses any kind of intrinsic value in a certain degree, then not only must that same thing possess it, under all circumstances, in the same degree, but also anything *exactly like it,* must, under all circumstances, possess it in exactly the same degree.[3]

1 In *An Abstract of A Treatise of Human Nature.*

2 See Jaegwon Kim, "Supervenience and Nomological Incommensurables," *American Philosophical Quarterly* 15 (1978): 149–56.

3 *Philosophical Studies* (London, 1922), p. 261.

Moore himself, however, seems not to have used the term "supervenience"; it was R. M. Hare, I believe, who, writing many years later, gave it the philosophical currency it now enjoys. The following passage from Hare is now generally recognized as a classic source that helped to shape the initial contours of the concept:

> First, let us take that characteristic of "good" which has been called its supervenience. Suppose that we say "St. Francis was a good man." It is logically impossible to say this and to maintain at the same time that there might have been another man placed exactly in the same circumstances as St. Francis, and who behaved in exactly the same way, but who differed from St. Francis in this respect only, that he was not a good man.[4]

Here, Hare speaks of supervenience as a "characteristic" of the term "good." However, it is clear that it is more usefully construed as a *relation,* a relation between "good" and terms that denote such things as patterns of behavior and traits of character. What Hare is saying is that it is "logically impossible" for there to be two persons who are exactly alike in these latter respects and yet differ in respect of being a good man. It is also clear that supervenience is better thought of as a relation not between properties or terms taken singly but between *sets* or *families* of them. Thus, we can say that all valuational properties (that is, the set of all valuational properties) are supervenient upon the set of all natural or descriptive properties. We shall in this paper discuss supervenience chiefly for properties rather than predicates; the choice here is indifferent to an extent, but, as we shall see, not wholly so. One could also speak of supervenience for sentences, facts, events, propositions, and languages; I shall argue below that fact supervenience can be understood in terms of property supervenience. It will become plausible, I believe, that property supervenience is fundamental, and that supervenience for most other entities can be explained in terms of it.

It is this evident generalizability beyond the sphere of ethics that makes supervenience an attractive and promising concept worthy of closer attention. Perhaps because of this, one now sees an increasing use of the term "supervenience" in a variety of areas, indicating the presence of substantial shared intuitive content. Thus, the aesthetic properties of a work of art have been claimed to be supervenient on its physical properties.[5] Some

4 *The Language of Morals* (London, 1952), p. 145.

5 Frank Sibley, "Aesthetic Concepts," *Philosophical Review* 68 (1959): 421–50; Jerrold Levinson, "Aesthetic Supervenience," *Southern Journal of Philosophy* 22, Supplement (1984): 93–110.

philosophers have found in psychophysical supervenience an attractive alternative to reductionist physicalism; it is thought that the supervenience thesis acknowledges the primacy of the physical without committing us to the stronger claims of physical reductionism.[6] The idea that valuational terms in general supervene on nonvaluational ones has been extended to epistemic terms, terms used for making epistemic appraisals, such as "evident," "certain," and "justified." The view that criteria of epistemic justification must be stated in nonepistemic terms can be thought of as an expression of the thesis that epistemic properties are supervenient on nonepistemic characteristics and relationships (e.g., causal properties and logical relations).[7] Leibniz's obscure doctrine of the dispensability of relational judgments is perhaps interpretable as the thesis that relations are supervenient on properties.[8] Quine's thesis of translational indeterminacy is usefully construed as the denial of the claim that meaning supervenes on the totality of physical fact.[9] Mereological supervenience has already been mentioned. There are other interesting questions we might formulate in terms of supervenience: Are causal relations supervenient on particular matters of noncausal fact? Are laws supervenient on their instances? Do theories supervene on data? Often the belief that there is a supervenience relation in a given domain, for example in the domain of the mental vis-à-vis the physical, forms an implicit premise of great importance that motivates and shapes the specific theories concerning that domain. Acceptance or rejection of the supervenience of the mental on the physical leads to the most basic division between theories of the mind–body relation: theories that accept psychophysical supervenience are fundamentally materialist, and those that reject it are fundamentally antimaterialist. This difference seems philosophically more basic and more

6 See note 11; also Kim, "Psychophysical Supervenience," Essay 10 of this volume, and "Psychophysical Supervenience as a Mind–Body Theory," *Brain and Cognition Theory* 5 (1982); Stephen P. Stich, "Autonomous Psychology and the Belief-Desire Thesis," *The Monist* 61 (1978): 573–91; John Haugeland, "Weak Supervenience," *American Philosophical Quarterly* 19 (1982): 93–103.

7 See Ernest Sosa, "The Foundations of Foundationalism," *Noûs* 14 (1980): 547–64, esp. p. 551. Also Alvin I. Goldman, "What is Justified Belief?" in *Justification and Knowledge*, ed. G. S. Pappas (Dordrecht, 1979), in which Goldman says he is looking for *nonepistemic conditions* for justified belief.

8 See, e.g., Hide Ishiguro, *Leibniz's Philosophy of Logic and Language* (Ithaca, 1972). It seems that Leibniz used the Latin word "supervenire" in stating his theory; see the quotation in footnote 3 on p. 71 of Ishiguro, op. cit.

9 This is especially clear in Quine's reply to Chomsky in *Words and Objections*, ed. D. Davidson and J. Hintikka (Dordrecht, 1969), esp. pp. 302 f.

significant than the usual classification of mind–body theories as "monist" or "dualist."[10]

This paper is intended as a general discussion of supervenience as a relation of dependency or determination. I shall be claiming that there are two separable concepts of supervenience, one stronger than the other, and that often what is offered in a philosophical discussion is the weaker of the two whereas what is needed is the stronger one. I shall also argue that the stronger relation is equivalent to "global supervenience,"[11] an alternative conception favored by some writers. One issue that will receive attention is what supervenience between two domains entails about the existence of kind-to-kind connections between them, and what this means for such relations as definability and reducibility between the two domains.

II. WEAK SUPERVENIENCE

The passage quoted above from Hare suggests this initial conception of the supervenience relation: the moral is supervenient on the natural in the sense that if two objects (persons, acts, states of affairs, and the like) are alike in all natural respects they must of necessity be alike in all moral respects. That is to say, things cannot differ with respect to some moral characteristic unless there is some natural property with respect to which they differ. Much the same idea is present in Donald Davidson's formulation of psychophysical supervenience:

> Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respects, or that an object cannot alter in some mental respects without altering in some physical respects.[12]

Here Davidson gives two explanations of supervenience, the first stated for events and the second for objects. I am not focusing on the fact that one is for events and the other for objects; I am only interested in the

10 For more details on this point, see Kim, "Psychophysical Supervenience as a Mind–Body Theory," op. cit.

11 (Added 1993.) This claim is controversial; see Essay 5 and "Postscripts on Supervenience," section 3, in this volume.

12 "Mental Events" in *Experience and Theory,* ed. L. Foster and J. W. Swanson (Amherst, 1979), p. 88.

general forms of the two explanations, and want to point to the fact that the first conforms to the pattern indicated in Hare's statement: mental characteristics supervene on physical ones in that no two things (objects, events, and the like) could differ with respect to some mental characteristic unless they differed also in some physical characteristic – that is, coincidence in the physical entails coincidence in the mental. If we were to create an exact physical replica of you, it and you would be psychologically indistinguishable. (Davidson's second explanation, as I shall suggest later, indicates a stronger relation of supervenience.)

A general analysis of supervenience that captures these ideas is straightforwardly developed. Let A and B be two nonempty families of properties (for simplicity we exclude relations) closed under the usual Boolean property-forming operations, complementation, conjunction, and disjunction (and perhaps others such as infinite conjunction and disjunction). This then is a definition of "weak supervenience" (the reason for calling it "weak" will be made clear below):

A *weakly supervenes* on B if and only if necessarily for any x and y if x and y share all properties in B then x and y share all properties in A – that is, indiscernibility with respect to B entails indiscernibility with respect to A.

We shall call A the *supervenient family* and B the *supervenience base* (*family*); properties in A are *supervenient properties,* and those in B are the *base properties.*

As an example: consider the set, A, containing the property of being a good man (G) and having the Boolean closure property; and let B be the set containing the property of being courageous (C), that of being benevolent (V), and that of being honest (H), and closed under the Boolean operations. A contains only two properties, G and $-G$, besides the tautological one ($G \vee -G$) and the impossible one ($G \& -G$). Suppose A weakly supervenes on B. This means that if two men share the same properties in B, say, both are honest and benevolent but lack courage (this will insure they share all other properties in B), then they must both be good men or neither is (they of course cannot differ in regard to the tautological or impossible property). Or, what is the same, if one is a good man but the other is not, there must be some property in B with respect to which they differ (say, the first is courageous but the second is not). Any differences in A must be accounted for by some difference in B.

To fix this further in mind: consider what we may call *B-maximal properties:* these are the strongest consistent properties constructible in B, and for our present example there are eight of these: $C \& V \& H$, $C \& V \&$

$-H$, $C \& -V \& H$, . . . , $-C \& -V \& -H$. These properties are mutually exclusive, and every object must have just one of these. Clearly, two objects are indiscernible in B just in case they have the same B-maximal property. Weak supervenience of A on B therefore comes to this: any two objects with the same B-maximal property must have the same properties in A – they are both G, or both $-G$. Or, using the terminology of "possible world," we may say: there is no possible world in which two objects have the same B-maximal property and yet differ in respect of G.

Given weak supervenience of A and B, therefore, within each possible world generalizations of the following form will hold:

(1) $(\forall x)[B_i(x) \rightarrow G^*(x)]$,

where, for each i, B_i is a B-maximal property and G* is either G or $-G$. Whether G or $-G$ is to be associated with a given B-maximal property is a feature of the specific possible world; but within each world these exceptionless universal conditionals between the property of being a good man on the one hand and the virtues of courage, benevolence, and honesty on the other must hold. Within each world, in fact, the following biconditionals hold:

(2) $(\forall x)[B^*(x) \leftrightarrow G(x)]$
$(\forall x)[B\#(x) \leftrightarrow -G(x)]$

where B* and B# are each a disjunction of B-maximal properties.

All of these points remain valid for weak supervenience generally, when B is finite; if B is not finite, these results will depend on the formability of B-maximal properties, which requires infinite conjunction, and their infinite disjunctions. I shall argue later that these operations are acceptable for properties (as distinguished from predicates), and they will be assumed in some of the formal arguments below (it will be clear exactly where they are used).

I dwell on these details in order to make the point that, although the definition of "weak supervenience" follows very closely the bench-mark explanations of supervenience in the literature, as witness the quotations from Hare and Davidson, the relation it defines is considerably weaker than one might have expected – indeed, too weak for some of its typical intended applications. The key to seeing this is that in a generalization of the form (1) above, which associates a supervenient property for each maximal property in the base family, whether G* is G or $-G$ depends on the particular world under consideration, and is not a feature invariant

across possible worlds. This means that weak supervenience of A and B (returning to our example) permits the following:

(a) In this world anyone who is courageous, benevolent, and honest is a good man, but in another possible world no such man is good; in fact, every such man is evil in this other world.
(b) Again, in this world anyone who has courage, benevolence, and honesty is good; in another world exactly like this one in respect of the distribution of these virtues, no man is good.
(c) In another possible world just like this one in respect of who has, or lacks, these traits of character, every man is good.

It is plain that weak supervenience permits these possibilities, for it only requires that *within* any possible world there not be two things agreeing in B but diverging in A, and this condition is met in each of these cases. It does not require that if in another world an object has the same B-properties that it has in this world, it must also have the same A-properties it has in this one. The particular associations between A-properties and B-properties in a given world cannot be counted on to carry over into other worlds.

Thus, weak supervenience falls short of the following condition: fixing the base properties of an object fixes its supervenient properties. This condition expresses a presumptive desideratum on the explication of supervenience: base properties must *determine* supervenient properties in the sense that once the former are fixed for an object, there is no freedom to vary the latter for that object. Weak supervenience goes some way toward this idea of determination: if you fix the base properties of two objects in the same way in a given world then you must fix their supervenient properties in the same way in that world. But under weak supervenience that is as far as the base properties constrain the attribution of the supervenient properties. That this is less than what we might expect of a relation of determination or dependence can be seen in various ways. Determination or dependence is naturally thought of as carrying a certain modal force: if being a good man is dependent on, or is determined by, certain traits of character, then having these traits must *insure* or *guarantee* being a good man (or lacking certain of these traits must insure that one not be a good man). The connection between these traits and being a good man must be more than a *de facto* coincidence that varies from world to world. We should be able to say: although Charles is not a good man, he *would* be one if only he *had* some benevolence in his nature as well as being honest and courageous. We should also be able to say: anyone who has these three virtues *would* be a good man although it is unfortunate that no

one has them all. Claims like these seem integral to what we mean when we speak of "good-making characteristics": any "X-making characteristic" must be such that if anything had it, it must of necessity have X (at least, it must necessarily be of positive relevance to its having X). Weak supervenience of moral upon nonmoral properties does not entail that there are nonmoral "conditions" or "criteria" for moral properties.

Another idea that is often associated with the idea of supervenience is this: if the moral supervenes on the nonmoral, any two worlds exactly alike in all nonmoral respects must be alike in all moral respects (in fact, they must be one and the same world). But this does not obtain under weak supervenience, as we have already seen. Similar points can be made about psychophysical supervenience: weak psychophysical supervenience is consistent with the existence of a world that is just like the actual one in every physical detail but in which no mentality, no consciousness, is manifested, and also a world that is just like ours except that a low-grade pain permeates every object everywhere. Thus, if we were to look, with Davidson, to supervenience for a relation of dependency for the mental vis-à-vis the physical, we would likely not find it in weak supervenience.

I find the following remarks by Moore instructive as well as surprising:

> I should never have thought of suggesting that goodness was "non-natural," unless I had supposed that it was "derivative" in the sense that, whenever a thing is good (in the sense in question) its goodness (in Mr. Broad's words) "depends on the presence of certain non-ethical characteristics" possessed by the thing in question: I have always supposed that it did so "depend," in the sense that, if a thing is good (in my sense), then that it is so *follows* from the fact that it possesses certain natural intrinsic properties, which are such that from the fact that it is good it does *not* follow conversely that it has those properties.[13]

We need not know exactly what Moore meant here by the term "follow" or "depend" to know that its force exceeds weak supervenience. For weak supervenience, as we have seen, only requires that any two things having the same natural properties must be either both good or both not good. This surely is not enough for saying that a thing's being good "follows" from its having the natural properties it has; weak supervenience, therefore, cannot explicate the notion of "dependence" Moore had in mind.

Does weak supervenience then have any useful philosophical applications? Although it is evidently not strong enough to serve as an analysis

13 In Moore's "A Reply to My Critics" in *The Philosophy of G. E. Moore*, ed. P. A. Schilpp (Chicago and Evanston, Illinois, 1942), p. 588.

of a full relation of dependence or determination, I believe it marks an interesting and significant relation of *partial* dependence or determination. Consider the case of moral supervenience again: perhaps all Hare wanted was weak supervenience.[14] Under weak supervenience there would be an inconsistency in one's commending an object (saying that it is good) but failing to commend another that is, or is believed to be, exactly like it in all descriptive details; however, there is nothing inconsistent, or incoherent, in failing to commend either while acknowledging the same descriptive properties of the two objects. This, in essence, is the prescription "Treat like cases alike" in ethical contexts. Weak supervenience, therefore, gives us the much discussed *Principle of Universalizability* of ethical judgments understood as a *consistency requirement.*[15] There is, however, a stronger sense in which the universalizability of ethical judgments has been understood: every singular ethical judgment must be supportable by a fully *general covering principle.* This stronger requirement goes beyond weak supervenience, corresponding rather to the notion of "strong supervenience" to be explained in the following section. That these two versions of the Universalizability Principle turn out to correspond nicely with the two concepts of supervenience distinguished in this paper speaks well for the naturalness as well as philosophical interest of the two concepts. These remarks about two Principles of Universalizability obviously apply to other cases involving valuational judgments (e.g., in aesthetics and epistemology).

Davidson has likened the relationship between the semantic notion of truth and syntactical concepts to psychophysical supervenience: in spite of the fact that truth is not definable or reducible in terms of syntax there is a sense in which the truth of a sentence depends on its syntactic properties.[16] This can, I think, be taken as something like weak supervenience: any two sentences that are syntactically indiscernible are in fact the same

14 According to Haugeland's report of a conversation with Hare (in Haugeland, op. cit.), it seems likely that what Hare had in mind is only my weak supervenience. This impression is confirmed by Professor Hare's inaugural address to the Aristotelian Society entitled "Supervenience," *Aristotelian Society,* Supplementary Volume 58 (1984): 1–16. The notions of "supervenience" and "entailment" as Hare explains them in his address turn out to correspond, roughly, to my weak and strong supervenience, respectively.

15 The distinction between two versions of the universalizability requirement is borrowed from J. Howard Sobel's unpublished notes on "Dependent Properties"; I also owe to Sobel the quotation from Moore's "Reply" (see note 13). Also useful in this connection are Monroe C. Beardsley, "On the Generality of Critical Reasons," *Journal of Philosophy* 59 (1962): 477–86; and Robert L. Holmes, "Descriptivism, Supervenience, and Universalizability," *Journal of Philosophy* 63 (1966): 113–19.

16 "Mental Events," p. 88.

sentence and must therefore have the same truth value. But obviously the truth value of a sentence cannot in general be relied on to be stable from world to world. Davidson's use of this example to explain supervenience points to the possibility that weak supervenience is also what he had in mind in speaking of psychophysical supervenience. This interpretation fits in neatly with Davidson's doctrine of psychophysical anomalism to the effect that there are no lawlike connections between mental and physical kinds. Lawlike connections must be stable over possible worlds (at least relative to some accessibility condition), and such connections between the mental and the physical are exactly what weak psychophysical supervenience does not require. On the other hand, this interpretation has a weakness: any robust materialist position should affirm, I think, that what is material determines all that there is in the world,[17] and this weak supervenience cannot give us. Although I am not sure whether Davidson would accept a full materialist position in my sense it seems that he wants more than weak supervenience.

Although it falls short of full-fledged materialism, weak psychophysical supervenience may be a possible thesis worth pondering: one might argue, for example, that although no physical fact about an organism, whether its behavior or its physiology, compels us to attribute to it some particular mental state, or any mental state at all, consistency requires that if two organisms manifest the same behavior and physiology, the same mental state must be attributed to each, and that this is the only constraint on the ascription of mental states. I think some such view may be held by those who take the attribution of mental states as just another case of positing theoretical explanatory states (relative to, say, behavior), and who take the possibility of the "inverted spectrum" seriously.

Another related case is this: even if, as many philosophers believe, theories are "underdetermined" by all possible data, they may be weakly supervenient on data in the following sense: although no set of data compels the choice of a particular explanatory theory, "relevantly similar data" must be explained by "relevantly similar theories." Weak supervenience as applied here thus yields a consistency requirement on theory construction in the same way that weak moral supervenience yields a consistency

17 For formulations of materialism see Terence Horgan, "Supervenience and Microphysics," *Pacific Philosophical Quarterly* 63 (1982): 29–43; Horgan, "Supervenience and Cosmic Hermeneutics," *Southern Journal of Philosophy* 22, Supplement (1984): 19–38. David Lewis, "New Work for a Theory of Universals," *Australasian Journal of Philosophy* 61 (1983): 343–77; for a related approach see Geoffrey Hellman and Frank Thompson, "Physicalism: Ontology, Determination and Reduction," *Journal of Philosophy* 72 (1975): 551–64, and "Physicalist Materialism," *Noûs* 11 (1977): 309–45.

requirement on ethical judgments, clarifying one precise way in which data constrain theory. There may be other interesting applications of weak supervenience; I hope, though, that what we have seen is enough to persuade us of its potential interest as a philosophical concept.

III. STRONG SUPERVENIENCE

A clue to an appropriate way of strengthening weak supervenience to obtain a stronger relation is seen when we consider the following equivalent formulation of weak supervenience:

A *weakly supervenes* on B if and only if necessarily for any property F in A, if an object x has F, then there exists a property G in B such that x has G, and if any y has G it has F.

Let us first see that the two definitions are equivalent. First, we show weak supervenience given by the earlier definition entails that newly defined. Assume that for some F in A, x has F. We need to show, for some G in B, that x has G, and that anything y with G has F. Let G be the B-maximal property of x (in any given world under consideration). Then trivially x has G. To show that anything y with G has B: suppose some y has G. Since both x and y have G and G is a B-maximal property, x and y share all properties in B. So by weak supervenience as first defined, x and y must share all properties in A. But F is in A and x has F. So y, too, must have F.

Second, to show that the second definition entails the first: assume x and y share all properties in B, and suppose they do not share all properties in A – that is, for some F in A, x has F but y does not. Since x has F, weak supervenience as defined in the second definition entails that for some G in B, x has G, and anything with G has F. By assumption, x and y share all properties in B; so y, too, has G, whence y has F, yielding a contradiction.

The key aspect of the second definition is its last clause, the requirement that any object having G also has F. The force of this clause is that *within* each world this G-F generalization must hold; it does not require that the G-F connection be stable across worlds. This suggests that in order to get a stronger supervenience relation that will insure the stability of connections between supervenient properties and their base properties,

we should try prefixing this clause with a suitable modal operator. It turns out that this yields what we want.[18]

This approach is also suggested by the second explanation of supervenience offered by Davidson in the passage quoted earlier; as will be recalled, the explanation was this: "an object cannot alter in some mental respect without altering in some physical respect." The modal force of "cannot" and reference to mental and physical "respects" strongly suggest that a proper way to understand what Davidson has in mind here is in terms of a connection between mental and physical characteristics that is constant over possible worlds. The last quoted passage from Moore on the "dependence" of goodness on natural properties, too, suggests a similar approach.

So let A and B be families of properties closed under Boolean operations as before:

A *strongly supervenes* on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and *necessarily* if any y has G, it has F.

To illustrate this, let us return to the example of being a good man and the three character traits of courage, benevolence, and honesty. The idea of strong supervenience comes to this: if St. Francis is a good man, there must be some combination of these virtues (say, honesty and benevolence) such that St. Francis has it, and anyone who has it *must* be a good man. This particular combination of the traits, however, need not be the only one in the base family that can "ground" being a good man; Socrates, too, is a good man, but the virtues that he has are courage and honesty rather than honesty and benevolence. Socrates is a good man in virtue of being courageous and honest while St. Francis is a good man in virtue of being honest and benevolent. Generally speaking, a supervenient property will have *alternative supervenience bases* – base properties that are each sufficient for the supervening property. If A strongly supervenes

18 In an earlier paper, "Supervenience and Nomological Incommensurables," op. cit., I said, incorrectly, that weak supervenience as defined by the first definition in this paper could be "equivalently defined" by a definition that in effect defines "strong supervenience" below. David Sanford's careful comments led me to see that this was a mistake, and that there in fact were two concepts of interest here. I was also helped by Barry Loewer who sent me his unpublished material on supervenience in which an essentially identical distinction is made. Others who have pointed out to me the failure of the claimed equivalence include Anthony Anderson and James Van Cleve (in his unpublished "Defining Supervenience"). I first made use of this distinction in "Psychophysical Supervenience as a Mind–Body Theory," op. cit.

on B, the B-maximal property of an object is a supervenience base for every A-property the object has. But a B-maximal property will often be stronger than is needed to serve as a base for a given A-property, and what is of interest would be a *minimal base* in the sense that any property weaker than it is not a supervenience base. (In contrast, B-maximal properties can be called "maximal bases.") If being a good man strongly supervenes on natural properties, any good man's maximal natural property (perhaps, a long conjunction of all his natural properties) would be a supervenience base for being a good man; however, this is obviously more than what we need (it would include the person's height, weight, date of birth, etc.) and would be less than perspicuous. On the other hand, the conjunctive property of being honest and benevolent may constitute a minimal base – a substantially more informative and more useful notion that justifies us in saying that this man is good *in virtue of* his honesty and benevolence, that his being good *consists in* in his having these traits of character, or that he is good *because* he is honest and benevolent.[19]

The modal term "necessarily" occurs twice in the definition of strong supervenience. It is neither possible nor desirable to specify in advance how necessity is to be understood here; an appropriate specification must depend on the particular supervenience thesis under consideration, and different readings of "necessarily" will yield different supervenience theses to consider. For example, if one is interested in the supervenience of moral upon natural characteristics, both occurrences of the term are perhaps best taken to signal logical or metaphysical necessity. For psychophysical supervenience it is possible to interpret the first occurrence as metaphysical necessity and the second as nomological necessity; it is also possible to interpret both as metaphysical, or both as nomological. In the case of mereological supervenience the most plausible construal may be that the first occurrence signifies metaphysical necessity and the second nomological or physical necessity. The main point is that different readings of the modal terms will generate different supervenience theses, and that this flexibility is a desirable feature of the definition as stated. We should, therefore, leave an exact interpretation of "necessarily" as a parameter to be fixed for particular cases of application.

The following relationship between the two concepts of supervenience is obvious:

19 Similar problems arise for the notion of a causal condition; often what is of interest is a minimal set of conditions sufficient for the effect, not just any sufficient set.

(3) Strong supervenience entails weak supervenience; weak supervenience does not entail strong supervenience.

The following is also obvious:

(4) Both supervenience relations are transitive, reflexive, and neither symmetric nor asymmetric.

In most cases of interest supervenience seems in fact asymmetric; for example, although many have claimed the supervenience of valuational on nonvaluational properties, it is apparent that the converse does not hold. Similarly, although psychophysical supervenience is an arguable view, it would be manifestly implausible to hold that the physical supervenes on the psychological. This asymmetry of supervenience may well be the core of the idea of asymmetric dependence we associate with the supervenience relation. For when we look at the relationship as specified in the definition between a strongly supervenient property and its base property, all that we have is that the base property entails the supervenient property. This alone does not warrant us to say that the supervening property is *dependent on,* or *determined by,* the base, or that an object has the supervening property *in virtue of* having the base property. These latter relations strongly hint at an asymmetric relation. We have learned from work on causation and causal modal logic the hard lesson that the idea of causal dependence or determination is not so easily or directly obtained from straightforward modal notions alone; the same in all likelihood is true of the idea of supervenient determination and dependence. Ideas of dependence and determination, whether causal, supervenient, or of other sorts, stubbornly resist capture in simpler and more transparent terms. The only possibly helpful suggestion I have is this: the asymmetric dependence of a supervenient property upon its base property may well derive from the asymmetric dependence of a comprehensive and integrated system of properties, of which it is an element, upon a similarly comprehensive and systematic family of base properties. Thus, the supposed dependence of, say, pain as a mental occurrence on some electrochemical processes of the nervous system may well be due to the asymmetric supervenient dependence of the whole family of mental phenomena on physical processes. This latter asymmetry, according to the present account, is simply the fact (if it is a fact) that the mental strongly supervenes, in the sense defined here, on the physical but not conversely. So what I am suggesting is a kind of holism: individual dependencies are grounded in the dependency between systems, not the other way around.

IV. GLOBAL SUPERVENIENCE AND KIND-TO-KIND CONNECTIONS

We now turn to another approach to analyzing supervenience, an approach favored by some writers on psychophysical supervenience and materialism.[20] This alternative approach speaks globally of "worlds" and "languages," and yields what may be called a concept of "global" or "world supervenience."[21] Thus, psychophysical supervenience has been explained by saying that worlds that are physically indiscernible are psychologically indiscernible (in fact, such worlds are one and the same). The supervenience of the moral on the nonmoral too could be explained in a similar way: there could not be two worlds that are indistinguishable in every nonmoral detail and yet differ in some moral respect. As will be recalled, we used such formulations in our discussion of the weakness of weak supervenience. Some might prefer this global approach to our own with the thought that by making explicit references to property-to-property connections between the supervenient family and its base (as in our second definition of weak supervenience and that of strong supervenience), our definitions beg an important question against those who invoke supervenience precisely because of its promise as a dependency relation free of commitment to property-to-property connections that smacks of discredited reductionism of various sorts.

Now, whether two worlds are discernible or indiscernible psychologically (or physically, etc.) is essentially a matter of how psychological properties are distributed over the individuals of the two worlds. If the worlds differ in respect of some *general* psychological fact, this must be reflected in some difference in the *singular* psychological facts they contain. Thus, to say that two worlds are psychologically discernible is tantamount to saying that for some psychological property P and an individual x, x has P in one but not in the other; to say that two worlds are psychologically indiscernible is to say that for every psychological property P and every individual x, x has P in one just in case x has P in the other.

Let A and B be sets of properties as before, and consider:

A *globally supervenes* on B just in case worlds that are indiscernible with respect to B ("B-indiscernible," for short) are also A-indiscernible.

20 See the papers by Horgan, Haugeland, and Lewis cited in notes 6 and 17.

21 I borrow the term "global supervenience" from Paul Teller in his "Relational Holism and Quantum Mechanics," *British Journal for the Philosophy of Science* 37 (1986): 71–81.

Discussion in section II has already shown that global supervenience is stronger than weak supervenience. Here is an argument to show that global supervenience is equivalent to strong supervenience. To show strong supervenience entails global supervenience: assume w_1 and w_2 are B-indiscernible but A-discernible. Then for some F in A and some x, F(x) in w_1 but $-F(x)$ in w_2. Let B* be the B-maximal property of x in w_1; then, by the strong supervenience of A on B, necessarily $(\forall y)[B^*(y) \rightarrow F(y)]$. Since w_2 is B-indiscernible from w_1, B*(x) in w_2. Hence, F(x) in w_2, yielding a contradiction. Next, to show the converse: suppose strong supervenience fails. Then, for some object x and property F in A such that F(x), if any G is in B and x has G, G fails to entail F. ("G entails F" is short for "Necessarily anything having G has F.") This is equivalent to saying that for this x and F, the B-maximal property of x does not entail F. Let w* be the actual world: in w* we have F(x) and B*(x). Consider another world w# that is just like w* in the distribution of B-properties over individuals; in particular, B*(x) in w#. However, since B* does not entail F, we can consistently suppose that $-F(x)$ in w#.[22] Thus, w* and w# are B-indiscernible but A-discernible; that is, A does not globally supervene on B. This completes the argument.

Global supervenience, therefore, is nothing but strong supervenience. The equivalence of these two concepts has a mutually reinforcing effect: the fact that two independently conceived notions turn out to be equivalent testifies to their naturalness and intuitive philosophical content. Moreover, it shows that in terms of commitment to kind-to-kind correlations there is no difference at all between global and strong supervenience; the thought that global supervenience is free from such commitments is a mistake.

What if we defined global supervenience for "facts"? We might have something like this: "Facts of kind P supervene on facts of kind Q just in case worlds that are identical in regard to facts of kind Q are identical in regard to facts of kind P." This formulation does not explicitly mention properties of individuals in either the analysandum or the analysans; however, it seems essentially equivalent to our formulation above in terms of properties. For what is it for two worlds to be "identical in regard to facts of kind P"? Think of worlds as certain maximal classes of facts; then for two worlds to be identical in regard to facts of kind P is for them to contain the same facts of kind P. The maximality condition on worlds as

22 (Added 1993.) This is an error. For further discussion of this issue, see Essay 5, section 2, and "Postscripts on Supervenience," section 3, in this volume.

classes of facts would presumably entail that two worlds contain the same facts of kind P if and only if they contain the same *singular facts* of kind P. A singular fact, I take it, is something of the form *a is F*, where a is an individual and F a property; and to say that the fact that a is F is a fact of kind P (say, a psychological fact) amounts, arguably, to saying that F is a property of kind P (say, a psychological property). It follows then that for two worlds to be identical in regard to facts of kind P is for the following to hold: for any property F of kind P and any x, x has F in one world if and only if x has F in the other. Thus, the notion of identity of worlds in regard to facts of kind P comes to the notion earlier explained of indiscernibility of worlds with respect to a set of properties of kind P. Moreover, it is by now evident that on the present construal of "facts" and of what it is for a fact to be "of kind P," talk of "properties of kind P" can in general replace talk of "facts of kind P" in discussions of supervenience, and, in particular, that supervenience of facts is reducible to supervenience of properties.[23]

What does the supervenience of A and B imply about the existence of correlations between the properties in the two families? Part of the answer is already clear from the definitions of weak and strong supervenience:

(4) If A weakly supervenes on B, then for every F in A there is a property G in B such that $(\forall x)[G(x) \rightarrow F(x)]$.

(5) If A strongly supervenes on B, then for every F in A there is a property G in B such that *necessarily* $(\forall x)[G(x) \rightarrow F(x)]$.

Our earlier discussion showed that if infinite conjunction and disjunction are assumed, (4) can be strengthened to:

(4a) If A weakly supervenes on B, then for each property F in A there is a property G in B such that $(\forall x)[G(x) \leftrightarrow F(x)]$, that is, each A-property has a coextension in B.

Under the same assumption, a companion result can be shown for (5) as well:

(5a) If A strongly supervenes on B, then for each property F in A there is a property G in B such that necessarily $(\forall x)[G(x) \leftrightarrow F(x)]$, that is, every A-property has a *necessary coextension* in B.

23 Supervenience for states and events, too, is reducible to property supervenience if they are construed as property exemplifications; for such a conception of states and events see "Events as Property Exemplifications," Essay 3 of this volume. If the alternative conception of events associated with Davidson (see, e.g., "The Individuation of Events" in Davidson, *Essays on Actions and Events* (Oxford, 1980)) is adopted, events could simply be treated as individuals, that is, as values of the variables "x," "y," etc., in the definitions of supervenience, again making a special notion of event supervenience unnecessary.

The following proves (5a): Let F be a property in A. We may assume F to be contingent; i.e., some x has F in some possible world w. ((5a) is trivially true for noncontingent F.) By the definition of strong supervenience there is a property G in B such that x has G (in w) and necessarily $(\forall y)[G(y) \rightarrow F(y)]$. Let $B_{x,w}$ be the B-maximal property x has in w. We have then:

$$\text{Necessarily } (\forall y)[B_{x,w}(y) \rightarrow G(y)],$$

whence:

$$\text{Necessarily } (\forall y)[B_{x,w}(y) \rightarrow F(y)].$$

And for each v that has F in a world u, we will have:

$$\text{Necessarily } (\forall y)[B_{v,u}(y) \rightarrow F(y)].$$

Let B* be the infinite disjunction of these B-maximal properties; then

$$\text{Necessarily } (\forall y)[B^*(y) \rightarrow F(y)].$$

It is easy to see we also have the converse:

$$\text{Necessarily } (\forall y)[F(y) \rightarrow B^*(y)].$$

For suppose not; then in some world w#, there is an object x such that F(x), but not B*(x). But by strong supervenience there is some property K in B such that K(x) in w# and necessarily $(\forall y)[K(y) \rightarrow F(y)]$. Let B# be the B-maximal property of x in w#. Then, as before, necessarily $(\forall y)[B\#(y) \rightarrow F(y)]$, and it follows that B# is one of the disjuncts in B*. Hence, x must have B*, yielding a contradiction. We thus have:

$$\text{Necessarily } (\forall y)[B^*(y) \leftrightarrow F(y)].$$

Note that the force of "necessarily" in this biconditional is that of the inner modal term (that is, the second occurrence of "necessarily") in the definition of strong supervenience. Depending on whether in a given case of supervenience we have logical, metaphysical, or nomological necessity for this term, we have in that sense of necessity a necessarily coextensive property in the base family for every supervenient property. In the case of nomological necessity, some might question this; we will take up this issue in the following section.

V. SOME PHILOSOPHICAL CONSIDERATIONS

The principal conclusions of the preceding section are, first, that strong supervenience is committed to the existence of a necessary coextension

in the base family for each supervenient property; and, second, that this commitment cannot be avoided by embracing global supervenience. For the two supervenience relations are in fact one. This should be found prima facie disturbing by some philosophers who have used supervenience to formulate certain philosophical claims. As we saw earlier, a full sense of dependency cannot be captured by weak supervenience; strong or global supervenience is needed. But to have this degree of dependency is *ipso facto* to be committed to the existence of a pervasive system of necessary property-to-property entailments, as is evident from the very definition of strong supervenience. And (5a) strengthens this: each supervenient property has a necessarily *coextensive* property in the base family. This may be more than what some philosophers thought they had bargained for.

We have already noted a possible dilemma in which Davidson may find himself: weak psychophysical supervenience appears too weak to yield materialism, but strong supervenience seems too strong in entailing the existence of a pervasive system of psychophysical equivalences. But Davidson's prime motive for advocating psychophysical supervenience is precisely to acknowledge the dependence of the mental on the physical but at the same time deny that there are laws connecting psychological and physical properties. What our results seem to show is this: if you want psychophysical dependence, you had better be prepared for psychophysical laws – or, at any rate, necessary psychophysical entailments. Some might dispute this line of thought on the ground that "nomological properties," i.e., those that are admissible in laws, are not closed under Boolean operations – that is, these operations, when applied to such properties, do not always yield properties fit to appear in laws.[24] B-maximal properties, their infinite disjunctions, and the like are "too complex," "too artificial and unnatural," and "too heterogeneous," it is argued, to be "natural kinds."

This raises various complex issues about the ontology and epistemology of laws, reduction, definition, and the like. I can, however, only indicate here the general approach I think we ought to take. First, we need to be sensitive to the distinction between *predicates* and *properties*, and beware that complexity or artificiality attaching to predicates (or linguistic

24 For interesting considerations along these lines see Paul Teller's "Comments on Kim's Paper," *Southern Journal of Philosophy* 22, Supplement (1984): 57–61. My remarks here, which include reactions to some of Teller's critical points, represent a modification and elaboration of the views I defended earlier, especially in "Supervenience and Nomological Incommensurables," op. cit.

constructions in general) need not attach to the properties they express. A long Boolean combination of predicates would normally be complex *qua* predicate; on the other hand, the property it expresses need not inherit that complexity (the Boolean expression may be equivalent to a short and simple one). The definitions in this paper have been framed for properties, not predicates; such operations as infinite conjunctions and infinite disjunctions would be highly questionable for predicates, but not necessarily for properties – any more than infinite unions and intersections are for classes. The property of being less than one meter long can be thought of as an infinite disjunction (e.g., of all properties of the form, being less than $n/n + 1$ meters long, for every natural number n). In fact, we could do with *sets* of properties, dispensing with infinite conjunctions and disjunctions. The main point is that there is no direct inference from the constructional details of properties to their complexity or artificiality, whatever these things may mean for properties.

When we speak of laws, we may have in mind either sentences or some nonlinguistic, nonconceptual, objective connections between properties. If laws are taken to be sentences, our results do not show that psychophysical supervenience entails the existence of biconditional laws. For we are given no guarantee that there are predicates, especially reasonably simple and perspicuous ones, to represent the constructed properties. Reformulating our basic definitions in terms of *predicates* rather than properties will not help; for that would make infinitary procedures highly dubious, perhaps unacceptable. Moreover, strong psychophysical supervenience stated for psychological and physical predicates seems considerably less plausible than when stated for properties: it asserts that for each psychological predicate there is a physical predicate that (logically or nomologically) entails it. What is the physical predicate that entails, say, "being bored"? It seems that we would at least need to appeal to "ideal physical languages" and the like to get started, and this might bring us right back to talk of properties.

If, on the other hand, laws are construed as objective connections between properties, which can be expressed by sentences or statements ("nomological" or "lawlike statements"), then (5a) must be accepted as stating that there are biconditional laws – if laws are "only" nomologically necessary, then equivalences that are at least as strong as laws – between supervenient and base properties. But what are the implications of this? Does it mean that the supervenient properties (or theories formulated in terms of them) are necessarily "reducible" to their bases? That moral properties are definable in naturalistic terms, that psychology is reducible

to physical theory, and so on? It might seem that these biconditional laws, or necessary equivalences, supply the "bridge laws" required by the classical conception of intertheoretic reduction. But this conclusion would be premature. Reduction, explanation, and the like are epistemic activities, and the mere fact that such equivalences or biconditionals "exist" is no guarantee that they are, or will ever become, *available* for reductive or explanatory uses. "Availability" here is best understood, I think, in terms of representation in a well-confirmed explanatory theory, and this in turn will depend on, among other things, our own cognitive powers, and our proclivities and idiosyncrasies in matters of what we find comfortable and satisfying as explanations. So the existence of a necessary physical coextension for every psychological property would have no direct bearing on our ability to carry out a physical reduction of psychology.

What we could more reasonably expect is this: as science makes progress, it will succeed in identifying an increasing number of *local* physical coextensions for psychological properties, that is, physical coextensions *restricted to specific domains* (e.g., particular biological species); and a sufficiently broad system of such local coextensions can serve as a base for "local reductions" of psychological theories. As many have pointed out, any given mental state is likely to have "multiple physical realizations" over distinct physical structures or biological species; however, for any given species or kind of structure, there may well be a uniform base, and if a comprehensive array of such bases is identified for, say, human psychological states, then human psychology could be "locally reduced" to physical theory.[25] If Martian psychological states, because of the different Martian anatomy and physiology, have different physical bases, Martian psychology would have to receive a different local physical reduction, even if the Martians and humans instantiate the same psychology.

Moore held the view that goodness is a "simple nonnatural" property, where by "simple" he meant indefinability and by "nonnatural" inaccessibility through normal sensory experience. Given this, it is somewhat remarkable that he was entirely unperturbed by the supervenience of goodness on natural properties. In fact, in the passage last quoted he says that *unless* he had thought that the goodness of an object "followed" from its natural properties he would "never have thought of suggesting that

25 On "local reductions" see Kim, "Psychophysical Supervenience as a Mind–Body Theory," op. cit.; also Robert C. Richardson, "Functionalism and Reductionism," *Philosophy of Science* 46 (1979): 533–58.

goodness was 'non-natural'." If the goodness of a thing "follows" from its natural properties which, we may assume, are accessible through normal sense-experience, why then isn't goodness itself so accessible? Perhaps, it might be replied that accessibility in this sense isn't the issue; what is crucial is that goodness has no direct or immediate "presentation" in sensory experience. But if goodness does follow from natural properties, why isn't this enough as a basis for a naturalistic epistemology of goodness, and why doesn't this make the intuitionist moral epistemology at best otiose?

It is interesting to note Moore's observation that, although the goodness of a thing follows from certain of its natural properties, from the fact that it is good it does *not* follow that it has these natural properties. Would our (5a) have discomfited Moore? Probably not, for the term "follow" as used by Moore, and perhaps also in its general philosophical usage, appears to have an unmistakable epistemological dimension: if goodness "follows" from certain natural properties, we should be able to "see" or "infer" that a thing is good by seeing that it has these natural properties. The necessary naturalistic coextension of goodness, as far as the arguments of this paper go, has no such epistemological status: we know it must exist, if strong supervenience obtains, but may never know "what it is." Nor can such a coextension be expected to provide a definitional basis for the term "good"; in fact, its existence does not suffice even to show the "in principle" definability of "good" in naturalistic terms. For the notion of definition carries certain semantic and epistemological associations, and even if we could identify the underlying naturalistic coextension of goodness we cannot expect these associations to hold for it.

We seem to have reached the conclusion that supervenience relations by themselves imply nothing directly about such relationships as definability and reducibility, if the possibility involved in "definability" or "reducibility" is construed in a fairly strong and realistic sense. However, as I shall argue later, there is a critical tension between acceptance of a supervenience thesis with regard to a pair of domains and rejection of all significant epistemic or conceptual relationships between the domains. But let us first briefly turn to the issue of *autonomy.*

What does supervenience imply about the autonomy of what supervenes in relation to its base? Although a thorough discussion would require a more precise understanding of the relevant concept of autonomy, it seems that weak supervenience can be entirely consistent with autonomy; that, in fact, may be one of its chief attractions. However, the case

is different with strong supervenience: under strong supervenience, the base wholly determines the supervening properties. If strong psychophysical supervenience holds, what happens in the realm of the mind is determined in every detail by what happens in the physical realm. This determinative relation is an objective matter; it does not depend on whether anyone knows anything about it, or what expressions are used to talk about mind and body. Unlike in the case of reduction and definition, epistemological considerations do not intrude here. That is perhaps why global supervenience is often used to state the doctrine of materialism. Likewise, strong supervenience of moral upon natural properties may signal a form of "moral naturalism" – not the definitional thesis of ethical naturalism, but a metaphysical thesis that recognizes the ontological primacy of the natural over the moral.[26]

Thinking about causal determination will, I believe, give us a useful point of analogy in thinking about supervenient determination. If causal determinism ("Every event has a cause") holds, every occurrence has a temporally earlier determinative condition. However, this says nothing about how successful we shall be in identifying causes and framing causal explanations; it is also silent on how successful we shall be in discovering causal laws. Explanation is an epistemological affair, and the claim that all events are causally explainable is an epistemological thesis, or a methodological doctrine, not entailed by the metaphysical thesis of universal causation alone (unless the former is expressly read so as to mean the latter). Similarly, the thesis that a given domain supervenes on another is a metaphysical thesis about an objectively existent dependency relation between the two domains; it says nothing about whether or how the details of the dependency relation will become known so as to enable us to formulate explanations, reductions, or definitions.

Having sundered the metaphysical thesis of causal determinism from its associated epistemological thesis about the possibility of causal explanation, we are now in a position to bring them together and appreciate their mutual relevance. There is, first, this much direct relationship: where there is no causal relation, there can be no correct causal explanation. When an event causes another, that constitutes the objective fact that makes a corresponding causal explanation "correct" or "true." More generally, we can think of the thesis of causal determinism as providing a

26 For a discussion of moral supervenience in relation to the problem of moral realism see S. W. Blackburn, "Moral Realism" in *Morality and Moral Reasoning*, ed. J. Casey (London, 1971).

metaphysical basis for the methodological strategy or principle enjoining us to search for explanations of natural events in terms of their causal antecedents, and also providing an explanation of why this strategy works as well as it does. Acceptance of causal determinism, therefore, can be viewed as an expression of a commitment to the method of causal explanation as an epistemological strategy. Conversely, it is our success, limited though it may be, in discovering causal connections and formulating causal explanations that forms an essential basis of our belief in causal determinism.

Similarly, the belief that a supervenience relation obtains for a pair of domains can motivate our search for specific property-to-property connections in terms of which illuminating reductions and edifying definitions might be formulated. Where strong supervenience obtains, (5a) gives us the assurance that such connections in the form of necessary equivalences are there to be discovered, without of course the further assurance that we shall succeed in discovering them or that they will be representable in an explanatory theory. A case in point is mereological supervenience, the doctrine that the macro-properties of material things are supervenient on their micro-properties. It is this metaphysical doctrine of atomism that seems to underlie and support the enormously productive research strategy of micro-reduction in modern theoretical science. And, conversely, the success of this research strategy reinforces our belief in mereological supervenience. Perhaps, similar remarks apply to moral supervenience: the belief that the moral supervenes on the nonmoral may have shaped some of the major assumptions and tasks of moral philosophy, such as the search for naturalistic definitions of ethical terms, the belief that there must be nonmoral "criteria" for moral ascriptions, the belief that there are such things as "good-making" or "right-making characteristics," and the perennial attempts to state conditions for the rationality of acts or justness of institutions in naturalistic and descriptive terms.

I think these remarks explain the point of (5a): it helps us to see the connection between the thesis of supervenience concerning a pair of domains and a certain epistemological strategy we may adopt in regard to them. It explains how the belief in the supervenience thesis can lead to, and in turn be supported by, the expectation that one domain can be understood – reduced, defined, explained, etc. – in terms of the other through the discovery of necessary equivalences that (5a) assures us must exist. The tension I alluded to earlier, several paragraphs back, arises pre-

cisely because this connection is contravened in embracing supervenience but rejecting at the same time all significant conceptual or epistemic relationships. That is to say, there is a sense, though a weak one, of possibility in which (5a) shows that strong supervenience entails the possibility of reduction or definition across the domains involved.[27]

27 In addition to the persons whose help has already been cited, I am indebted to David Benfield, Earl Conee, Fred Feldman, John Heil, Terence Horgan, Arnold Koslow, Brian McLaughlin, Robert Richardson, Ernest Sosa, and Paul Teller.

5

"Strong" and "global" supervenience revisited

In an earlier paper, "Concepts of Supervenience,"[1] I characterized two distinct concepts of supervenience, "strong" and "weak," and compared them with each other and with a third concept, "global supervenience." In this paper I wish to correct an error in the earlier paper and present further material on supervenience, including a new characterization of strong supervenience, which I believe is particularly perspicuous, and a discussion of the adequacy of global supervenience as a determination relation. I shall also present a strengthened relation of global supervenience based on *similarity* rather than *indiscernibility* between worlds, which may well be a more useful concept than the currently popular conception of global supervenience.

1. A NEW CHARACTERIZATION OF "STRONG SUPERVENIENCE"

Let A and B be two sets of properties (closed under complementation, conjunction, disjunction, and perhaps other property-forming operations).[2] A is said to *weakly supervene* on B just in case:

(I) Necessarily, for any x and y, if x and y share all properties in B, then x and y share all properties in A – that is, indiscernibility in B entails indiscernibility in A.

This corresponds in a straightforward way to the informal characterization of supervenience commonly found in the literature.[3] As was shown in the

I have received helpful comments on topics discussed here from David Lewis, Barry Loewer, Joe Mendola, Brad Petrie, Ernest Sosa, Paul Teller, Nicholas White, and Stephen Yablo.

1 Essay 4 of this volume.

2 Infinite conjunction and disjunction are needed for some of the arguments in "Concepts of Supervenience."

3 See, e.g., R. M. Hare, *The Language of Morals* (London: Oxford University Press, 1952), p. 145.

earlier paper, weak supervenience can be equivalently explained as follows:

(II) Necessarily, for any object x and any property F in A, if x has F, then there exists a property G in B such that x has G, and if any y has G, it has F.

As I argued, although supervenience thus characterized (especially by (I)) corresponds, essentially word for word, to the notion that many philosophers profess to have in mind, it yields only a fairly weak relationship seemingly inadequate for many of the uses to which it has been put; in particular, it seems too weak to fully capture the intuitive relation of "determination" or "dependency" between sets of properties (or facts, states, and the like). The following stronger relation of supervenience, therefore, was introduced: A is said to *strongly supervene* on B just in case:

(III) Necessarily, for any object x and any property F in A, if x has F, then there exists a property G in B such that x has G, and *necessarily* if any y has G, it has F.

This definition closely parallels the second characterization above, (II), of weak supervenience, the only difference being the presence of the modal expression "necessarily" governing the general conditional "if any y has G, it has F." The insertion of this modal qualifier guarantees world-to-world stability for the correlations between supervenient properties and their "base properties," a feature conspicuously lacking in weak supervenience.

One question not addressed in "Concepts of Supervenience" is whether there is an alternative definition of strong supervenience which formally parallels the first characterization (I) of weak supervenience in the same way that the definition (III) of strong supervenience parallels the second definition (II) of weak supervenience. Such a definition would be useful and instructive since (I) is the most common ("canonical," one might say) way of explaining supervenience, perhaps because of its appealingly intuitive and perspicuous form. A characterization of strong supervenience with a similar form would have similar virtues, wearing on its sleeve the nature of its relationship to weak supervenience. Of course the definition, (III), helps in a limited way, when contrasted with the second definition, (II), of weak supervenience. The limitation stems from the fact that the second definition of weak supervenience, although formally equivalent to the first, lacks its familiarity and intuitive clarity.

In a forthcoming paper,[4] Brian McLaughlin introduces a relation of supervenience (of A on B) characterized thus:

(IV) For any worlds w_j and w_k, and for any objects x and y, if x has in w_j the same B-properties that y has in w_k, then x has in w_j the same A-properties that y has in w_k.

McLaughlin's definition, as it turns out, is what we are looking for: the relation it characterizes is equivalent to strong supervenience as given by (III) and it is obvious that McLaughlin's definition formally parallels definition (I) of weak supervenience. It differs from (I) only in that indiscernibility in A or B can be "cross-world" as well as within a single world. Thus, we might put McLaughlin's definition like this: A *strongly supervenes* on B just in case *cross-world indiscernibility in* B *entails cross-world indiscernibility in* A.

Another virtue of the new definition is this: What makes weak supervenience weak is, as noted, the fact that it does not guarantee stability across worlds for the correlations between supervenient properties and their base properties – that is to say, these correlations have no modal force. However, the earlier definition, (III), introduces this guarantee through an explicit clause, an easy but, for that reason, somewhat uninstructive way of doing things. A better way would have been to construct a definition of strong supervenience that has a certain intuitive rightness in its own right and from which one could show, by an argument, that supervenience under this conception turns out to have the desired characteristics. McLaughlin's definition clearly fills the bill.

McLaughlin's supervenience is easily seen to be equivalent to strong supervenience as defined by (III). To show, first, that McLaughlin's supervenience entails strong supervenience: Assume, for any property F in A, x has F at w_i. Let B_i be the B-maximal property of x at w_i (the B-maximal property of an object at a world is the strongest B-property the object has in that world; it entails every B-property of the object; see "Concepts of Supervenience," Essay 4, section II). Let B_i be the G in the definition of strong supervenience; we need to show that necessarily if any y has it, it has F. Suppose otherwise – that is, at some w_j there is a y such that y has B_i but not F. Thus, x has B_i at w_i and y has B_i at w_j; that is, x and y have the same B-properties in these worlds respectively, and by McLaughlin's supervenience, they must have the same A-properties in the respective

4 "Why Try to Bake an Intentional Cake With Physical Yeast and Flour?" (unpublished manuscript).

worlds. Since x has F in w_i, y must have F in w_j, contradicting the supposition. Hence, A strongly supervenes on B.

To show the converse: Assume x at w_i and y at w_j share the same B-properties. Let F be any A-property that x has at w_i; we need to show that y has F at w_j. By strong supervenience there is a property G in B such that x has G at w_i, and necessarily (that is, at every world) anything with G has F. Ex hypothesi, y at w_j has all the B-properties x has at w_i. So y has G at w_j, from which it follows that y has F at w_j. Hence, A supervenes on B in the sense of McLaughlin.

2. GLOBAL SUPERVENIENCE DOES NOT ENTAIL STRONG SUPERVENIENCE

In "Concepts of Supervenience," I claimed that strong supervenience is equivalent to "global supervenience" explained as follows:

> Any two worlds indiscernible with respect to B-properties are indiscernible with respect to A-properties.[5]

But this was an error, as Geoffrey Hellman, John Bacon, Bradford Petrie, and others[6] have pointed out. Strong supervenience does entail global supervenience; however, the converse does not hold, as Petrie shows with the following simple example:

> Consider the two worlds, w_1 and w_2, each with two individuals a and b. In w_1, a has G and F, and b has G. In w_2, a has G but not F, and b lacks G.[7]

It is clear that this pair of worlds is a counterexample to the strong supervenience of F on G (or, more verbosely, the unit set of F on that of G). However, the example leaves open the possibility that F globally supervenes on G: since the two worlds are not G-indiscernible they cannot be

5 Before this can be taken as a precise definition it must be sharpened in various respects; e.g., how indiscernibility with respect to a set of properties is to be understood for worlds not sharing the same domain of individuals, and whether or not "haecceitism" is accepted for individuals. These issues, however, will not affect the present discussion.

6 Hellman in personal communication and in "Determination and Logical Truth," *Journal of Philosophy* 82 (1985): 607–16 (see footnote 3, pp. 608–09); Bacon, "Supervenience, Necessary Coextensions, and Reducibility," *Philosophical Studies* 49 (1986): 163–76; Petrie, "Global Supervenience and Reduction," *Philosophy and Phenomenological Research* 48 (1987): 119–30. Paul Teller and Neil Tennant also called this to my attention.

7 "Global Supervenience and Reduction." Petrie first presented this counterexample in his doctoral dissertation, *Semantics and Physicalism*, The University of Michigan, 1985.

a counterexample to the global supervenience of F on G. Therefore, global supervenience cannot entail strong supervenience.[8]

3. IS GLOBAL SUPERVENIENCE STRONG ENOUGH?

But what is the metaphysical significance of the failure of global supervenience to entail strong supervenience?[9] To see Petrie's example as showing this failure is to see, I think, the limitation of global supervenience as a relation of determination or dependence. Given the existence of worlds like w_1 and w_2, which is permitted by global supervenience of F on G, does it make sense to think of F as being "dependent on," or "determined by," G? In w_1 a has G and also F; in w_2, however, a has G but lacks F; and the only fact in this world that might account for this failure of the G-F connection in a is the fact that another object b, which, we may assume,[10] is totally unrelated to a, has non-G, although it has G in w_1. Clearly it is possible to think of other bizarre pairs of worlds that violate strong supervenience but are compatible with global supervenience.

It is difficult to see how, given worlds like these, F's global superve-

choose. Moreover, as we saw, global psychophysical supervenience is consistent with there being within a given world, perhaps this one, two physically indistinguishable organisms with radically different psychological attributes. It is doubtful that many materialists would regard these consequences as compatible with their materialist tenets; it seems clear that they are not compatible with the claim that the mental is determined wholly by the physical.

Moreover, global supervenience without strong supervenience is difficult to understand. If the mental globally supervenes on the physical, that cannot be a brute and unexplainable fact, something we would want to accept as a fundamental, primitive fact about the world. We would feel, I think, that there should be an explanation of it. If, as strong supervenience affirms, there should exist appropriate connections between specific psychological properties of objects and their physical properties, that would give us a basis for an explanation. When we ponder the kinds of situations in which strong supervenience but not global supervenience fails, the failure of strong supervenience seems to make global supervenience unexplainable and incomprehensible. All this makes one wonder if there could ever be *evidence* of the sort we could reasonably expect to obtain that would support the global physical supervenience of the mental while ruling out its strong supervenience.[13] There is a strong inclination, I think, to look for an explanation of "global determination" in terms of specific "local determinations"; we would find global determination without local determination mysterious and difficult to understand. Perhaps, this is a manifestation of our micro-reductive proclivities. But then so much the worse for global supervenience without strong supervenience.

Is there anything positive to say in favor of global supervenience? In "Global Supervenience and Reduction," Petrie argues that it is especially well suited for the formulation of materialism in view of the well-known examples of Hilary Putnam and Tyler Burge[14] which appear to show that the contents of certain propositional attitudes can depend on factors external to the subjects to whom the attitudes are attributed. To adapt an example of Putnam's, consider Oscar and Oscar's counterpart on Twin

13 This point is recognized by Petrie, though apparently for somewhat different reasons, in "Global Supervenience and Reduction," Section 4. I believe he underestimates its seriousness.

14 Putnam, "The Meaning of 'Meaning,'" in his *Philosophical Papers,* Volume 2 (Cambridge: Cambridge University Press, 1975); Burge, "Individualism and the Mental," *Midwest Studies in Philosophy* 4 (1979): 73–121.

earth, which is just like the earth except only that on Twin Earth water is replaced everywhere by an observationally indistinguishable compound XYZ. We may assume Oscar and T. E. Oscar are molecule-for-molecule duplicates of each other.[15] It seems plausible to think that Oscar and T. E. Oscar, despite their physical indistinguishability, can have different beliefs: for example, Oscar believes that oil and water don't mix, while T. E. Oscar believes that oil and XYZ don't mix. This seems to show that beliefs do not in general strongly supervene on the physical states of the persons to whom they are attributed. Thus, it may be thought, we need a form of physicalism that is consistent with the failure of local determination of the mental by the physical, and global psychophysical supervenience seems precisely what we want. For it affirms that the psychological states of the world, taken as a whole, are determined by its physical states taken as a whole, without requiring every psychological state of an individual to be determined by *its* physical states.

How plausible is this argument? Without getting ourselves entangled in the dispute about "wide" and "narrow" content and related issues, we can see, I think, that this argument does not go as far as it might appear at first blush. Consider those causal-historical *relations* Oscar has had with respect to water, and more generally to his environment, *in virtue of* which his belief, but not T. E. Oscar's belief, has the content "oil and water don't mix," and similarly the causal-historical relations characterizing T. E. Oscar with respect to XYZ that account for *his* belief content. These relational properties, some of which are highlighted in Putnam's discussion, are the key to seeing how local determination can work here. That is, it is perfectly possible to construe these belief contents to be determined by, and strongly supervenient on, the physical properties, *including relational historical properties,* of Oscar and T. E. Oscar.[16] Further, it is *necessary* to construe the matter this way; for otherwise it would be entirely inexplicable why Oscar's belief has the content it has and not the content that T. E. Oscar's belief has, and vice versa. It is no accident that Oscar has beliefs about water, not about XYZ, and T. E. Oscar has beliefs about XYZ, not about water; we expect this difference to be grounded in certain further specifiable relevant differences between them – differences that are generalizable. And our expectation is, on the whole, not disappointed. Imagine what our reaction would be if we were not able to find

15 I think we can ignore the difficulty with this example that if water is everywhere replaced by XYZ the two Oscars could not be molecule-for-molecule indiscernible.

16 For other similar examples see my "Psychophysical Supervenience," Essay 10 of this volume.

any such difference because, we are told, *there was none*. That would be a situation in which we were asked to believe that there just was no intelligible basis for our ascribing different belief contents to Oscar and T. E. Oscar – that is, a situation in which belief attributions would become wholly mysterious and lose their sense. It is possible, of course, that the differences that ground the difference in content ascriptions include psychological features; but this would only mean that they, too, must in turn physically supervene in an appropriate sense if psychophysical supervenience is to hold.

Petrie also mentions the economic value of a coin as a property that is plausibly determined by the physical features of the world but not by those of the coin itself. That seems right, but only if we limit ourselves to the "intrinsic" physical properties of the coin. For the value of a coin is best taken as a relational property of the coin regarded as a physical object, a property something has only in its relation to a specific economic community. Under physicalism it would be plausible to regard this property as being determined by, and strongly supervenient on, the physical properties, again including relational ones, of the coin. Or consider the property of being the tallest man.[17] Whether I have this property depends not only on my height but on who else exists in this world. The property of being taller than any other man is a relational property, a property that an object has in virtue of its relationship to other individuals, and we of course expect its supervenience base to include the relational properties of individuals. If you have this property, it will strongly supervene on your other properties including relational ones (say, your height being eight feet and everyone else's height being less than that).

Examples like these, therefore, do not call for global supervenience. What they call for is an explicit recognition of relations and relational properties. The lesson we learn from the Putman and Burge-style cases that is relevant to the topic of supervenience is that contrary to what used to be taken for granted, many intentional states turn out not to be "intrinsic" or "internal"[18] to the subjects to whom they are attributed. Rather, they turn out to be "extrinsic" and "noninternal," dependent in complex ways on physical and social factors outside the subjects. It is no surprise then that we must seek a wider physical supervenience base for them, including relations and relational properties, if they are thought to be

17 Both Petrie and Teller have suggested examples of this kind.

18 For discussions of "internal" or "intrinsic properties" see my "Psychophysical Supervenience," Essay 10 of this volume; Lewis, "New Work for a Theory of Universals," and "Extrinsic Properties," *Philosophical Studies* 44 (1983): 197–200.

supervenient on the physical. A full account of these cases will be facilitated by a generalization of "strong" and "weak" supervenience to accommodate relations as well as properties.[19]

5. GLOBAL SUPERVENIENCE STRENGTHENED: SIMILARITY VS. INDISCERNIBILITY

There seems to be a natural way of strengthening the concept of global supervenience to blunt, at least to a degree, the thrust of the observations in the preceding section which, as we saw, appear to undermine the claim of global supervenience as a dependency relation. It makes sense to think that if the mental is dependent on the physical, then not only must any two worlds that are *physically indiscernible* be *psychologically indiscernible,* but also any two worlds that are *physically pretty much the same* must be *pretty much the same psychologically* as well – that is, worlds that are highly similar in physical respects must not show large psychological differences. This idea suggests the following concept of *similarity-based* global supervenience (of A on B):

(V) The degree to which any two worlds are similar in respect of B-properties is matched by the degree to which they are similar in respect of A-properties.

This definition is perhaps too strong: we may not want to require that two worlds that are B-*dissimilar* must also be equally A-*dissimilar.* For example, we may want to allow, under global psychophysical supervenience, the possibility that two worlds that are quite dissimilar in physical respects could display similar psychological characteristics (consider the "multiple physical realizability" of psychological states often invoked in discussions of the mind–body problem).[20] I think this is a debatable issue, but there is also the following weaker relation which will suffice for the present:

(VI) Worlds that are pretty much alike in B-properties are pretty much alike in A-properties.

If we understand indiscernibility as a limiting case of similarity where the degree of similarity is maximal, similarity-based global supervenience, under either of the two conceptions above, can be considered to entail the usual indiscernibility-based global supervenience as a special case.[21]

19 See "Postscripts on Supervenience," section 1, in this volume.

20 See Hilary Putnam, "Psychological Predicates," in *Art, Mind, and Religion,* ed. W. H. Capitan and D. D. Merrill (Pittsburgh: University of Pittsburgh Press, 1967).

21 Note that the similarity approach can be applied also to weak and strong supervenience, especially as these relations are characterized by (I) and (IV) in section I above.

It would be difficult to formulate general criteria for measurement of similarity in the sense intended here. Similarity in this sense can, and must, be evaluated along many seemingly incommensurable dimensions, and judgments of similarity are certain to be highly sensitive to a variety of contextual factors, giving rise to unresolvable disagreements. But these difficulties do not negate the substantial intuitive content this notion has for us; for it is only a generalization of the familiar notion of similarity with respect to a single property or determinable (e.g.; color, shape, etc.). It seems in any case at least as clear and robust as the related, and more general, notion of "comparative overall similarity" for worlds that David Lewis and others have exploited in developing the semantics of counterfactuals.[22] It surely is not more vague or problematic than the latter.

In any event, the strengthened relation of global supervenience requires that two worlds that are pretty much alike in the base properties must be pretty much alike in the supervenient properties. Thus, if the mental globally supervenes in this sense on the physical, a world that differs only minimally from this world in physical respects (in it Saturn's rings have one more ammonia molecule) cannot show large psychological differences (plants, but no creatures with brains, are conscious); perhaps it shouldn't show any psychological difference at all.

Consider again Petrie's two worlds which, while defeating the strong supervenience of F on G, were seen to be consistent with F's global supervenience on G. It is not obvious, however, that they are consistent with F's similarity-based supervenience on G. As may be recalled, in w_1 a has G and also F, and b has G; in w_2, a has G but not F, and b has non-G. Given these two worlds, does F globally supervene, in the new strengthened sense, on G? The answer depends on two things: How similar are the two worlds in regard to G, and how similar are they in regard to F? Clearly it is not possible to answer these questions for schematic examples; we would need to know, first of all, what properties F and G actually are, and if we are to make meaningful comparisons we would need to look at richer worlds, with more individuals and properties, and have some idea of what laws and regularities hold in them.

It is clear, however, that examples can be constructed that are consistent with similarity-based global supervenience, under either of the two definitions above, but not with strong supervenience; all we need to do is to think of two worlds that, while differing minimally from each other both

22 See Lewis, *Counterfactuals* (Cambridge: Harvard University Press, 1973), especially chap. 4.

in physical and in psychological respects, contains an object which, although it has identical physical properties in the two worlds, has negligibly different psychological properties. This will yield similarity-based psychophysical global supervenience. However, for any psychological property M the object has in one world but lacks in the other, there are no world-invariant general conditionals of the form "Anything with P has M," where P is a physical property of the object. So similarity-based global supervenience does not entail strong supervenience. And, unlike indiscernibility-based global supervenience, it is not entailed by strong supervenience either;[23] for we can imagine that mental characteristics, though strictly correlated with neural states, are in general critically sensitive to minuscule physical differences.

23 As Barry Loewer pointed out to me.