Introduction

ELIAS E. SAVELLOS AND ÜMIT D. YALÇIN

1. The Basic Concept

For most people who are not familiar with its many manifestations, analytical philosophy is the philosophy of reductionism par excellence. And the title is well earned when one recalls the string of reductionist programs that have left their mark on the first part of this century, ranging from the purported *analytical* reductions proposed by phenomenalism and behaviorism, to the weaker *theoretical* reductions of the later generations. Yet, starting with sporadic suggestions in the 1960s and 1970s, the philosophical literature is now rife with pronouncements of the wrongheadedness of all reductive programs. Perhaps surprisingly, the current literature associated with analytical philosophy is being swept by a wave of antireductionism.

Reductionism might be dead or dying, but the idea that certain entities we seem to talk and think about depend on others for their existence (and that they are somehow less real?) is still alive and kicking. This had led philosophers to search for a topic-neutral *nonreductive* dependence relationship that can be easily incorporated into the analytical toolbox of a variety of philosophical endeavors, performing at least part of the function reductive relationships were supposed to fulfill. Hence the recent philosophical interest in *supervenience,* which purports to be precisely this sort of relationship. Although the concept that the modern use of 'supervenience' aims to express has been around for some time, widespread interest in it is a relatively recent phenomenon.[1]

The basic idea behind the philosophical concept of supervenience is perhaps best introduced by an example. Take the property of being a beautiful piece of music.[2] When we consider the various musical works that instantiate this property, it is highly dubious that we can come up with any one sequence of musical notes or sounds[3] that is common to all and only the pieces that instantiate it. With the terminology that is quite commonly utilized in such circumstances, one could express this point by saying that the beauty of a musical piece seems to be *multiply realizable*. Hence, it is highly improbable that the

property in question can be correlated with any such sequence of notes or sounds. This is enough to convince most of us that being a beautiful piece of music cannot be *identified with* or be *reduced to* having any given sequence of sounds as a part.[4] But at the same time, we are still convinced that the beauty of any musical piece has to do with the sequence of sounds that constitutes it. If a musical piece is beautiful and some other piece is not, they cannot be constituted of exactly the same sequence of sounds; their aesthetic difference has to be somehow due to some difference in the sequence of sounds that constitutes each piece. The beauty of a musical piece seems to be *grounded in* the sequence of sounds that constitutes a musical work without being *identifiable* with any unique property of such sequences of sounds. The concept of supervenience is supposed to denote this dependence relationship that appears to be weaker than reducibility.

Hence, whether one talks about the supervenience of concepts, properties, phenomena, entities, or what have you, the basic notion of *A*s supervening on *B*s appears to subsume (a) *covariance*, where variations in *A*s are correlated with variations in *B*s, (b) *dependence* of *A*s on *B*s (or, if these are different,[5] the *determination* of *A*s by *B*s), and (c) the *nonreducibility* of *A*s to *B*s.

Given the character of the example we used, it might sound odd at this point to characterize supervenience as incorporating nonreducibility but not *irreducibility*. For it was the apparent irreducibility of the beauty of a musical piece that led us to look for a weaker dependence relationship. As we shall see later, the same goes for other fields in philosophy: it was the apparent irreducibility of mental properties to physical properties in the philosophy of mind, and of moral properties to natural properties in ethics, that opened the door to supervenience. Nevertheless, it has become common practice these days to explicate the notion of supervenience only in terms of nonreducibility,[6] and we shall stay with this perhaps controversial practice. This is undoubtedly an important issue that needs more attention, since the stand one takes will have a bearing on how one attempts to provide a more detailed explication of supervenience and the philosophical work one expects from this notion. Beyond this point, the issues become even more controversial.

2. The Logic of Supervenience

Mainly due to Jaegwon Kim's pioneering work on the topic, the prevalent understanding of supervenience is that of a relation between families of properties closed under a set of property-forming operations. That is, a set of properties, *A*, supervenes on another set of properties, *B*, if and only if there is a certain kind of relationship between each

A-property and a property that can be constructed from the *B*-properties by means of a set of specified property-forming operations. Yet, even if this is the intuitive understanding of supervenience, there still remains the task of clearly formulating specific theses that capture it. By the time the supervenience literature had ripened, there were at least five supervenience theses (more accurately, schemata) that purportedly expressed a relationship that could ground the ontic priority claims supervenience was supposed to support.[7] We give these theses below together with their symbolic representations, following Kim (1990a) in calling them "covariance" theses for reasons that will become clear in the further course of our discussion. Where A ranges over the properties in the supervenient set of properties, and B over the subvenient or base set of properties, and $=_x$ is used to indicate indiscernibility with respect to X-ness, we have the following:

(WC$_1$) Weak Covariance 1 For any possible world w and objects x and y in w, if x and y are indiscernible with respect to properties in B, they are indiscernible in respect of properties in A.

$\forall w\ \forall x\ \forall y\ ((x \text{ is in } w \text{ and } y \text{ is in } w) \rightarrow (\forall B\ (Bx \leftrightarrow By) \rightarrow\ \forall A\ (Ax \leftrightarrow Ay)))$

(WC$_2$) Weak Covariance 2 Necessarily, if anything has some property A_i in A, there exists a property B_j in B such that the thing has B_j, and everything that has B_j has A_i.

$\Box\ \forall x\ \forall A\ (Ax \rightarrow \exists B\ (Bx\ \&\ \forall y\ (By \rightarrow Ay)))$

(SC$_1$) Strong Covariance 1 For any objects x and y and worlds w_1 and w_2, if x in w_1 is B-indiscernible from y in w_2 (i.e., x has in w_1 precisely those B-properties that y has in w_2), then x in w_1 is A-indiscernible from y in w_2.

$\forall w_1\ \forall w_2\ \forall x\ \forall y\ ((x \text{ is in } w_1 \text{ and } y \text{ is in } w_2) \rightarrow (\forall B(Bx \leftrightarrow By) \rightarrow \forall A\ (Ax \leftrightarrow Ay)))$

(SC$_2$) Strong Covariance 2 Necessarily, if anything has some property A_i in A, there exists a property B_j in B such that the thing has B_j, and *necessarily*, everything that has B_j has A_i.

$\Box\ \forall x\ \forall A\ (Ax \rightarrow \exists B\ (Bx\ \&\ \Box\ \forall y\ (By \rightarrow Ay)))$

(GC) Global Covariance For any worlds w_1 and w_2, if w_1 and w_2 are B-indiscernible, then they are A-indiscernible.[8]

$\forall w_1\ \forall w_2\ (w_1 =_B w_2 \rightarrow w_1 =_A w_2)$

These should be understood as schemata that can be particularized in various ways so as to be a component of this or that supervenience thesis. A particularization will consist of a specific choice for the supervenient and subvenient properties, and a specific construal of the necessity operators (or alternatively, the range of the world-binding quantifiers). Whether any interesting thesis can be formulated by assigning different interpretations to different modal operators oc-

curring in the same schema (or by using different ranges for different world-binding quantifiers in the same schema) is an intriguing, yet not much discussed question.

Weak supervenience is so called because, roughly, it only requires *B*-indiscernibles in the *same* world to be *A*-indiscernible. Strong supervenience is stronger in that it also requires *B*-indiscernibles *across* possible worlds to be *A*-indiscernible. To put it in terms of a concrete example, if entities in a given world that had indiscernible brain states had to be in indiscernible mental states, weak supervenience would obtain. Further questions such as If there were other entities with brain states indiscernible from the entities in that world, what would their mental states be like? or Could there have been other entities with brain states indiscernible from the entities in that world, but with different mental states? are irrelevant to a weak supervenience claim, but not to a strong supervenience claim.[9] Finally, unlike both strong and weak supervenience, global supervenience characteristically takes whole worlds as the relata of the supervenience relation.

Given this proliferation of theses, one is naturally led to wonder which (if any) of these formulations is the most suitable. This, in turn, raises questions about the relationship of these various theses to each other. A significant portion of the supervenience literature, including some of the essays in this volume, is devoted to settling such questions.

With the foregoing characterizations, it is fairly certain that particularizations of (SC_1) entail comparable particularizations of (WC_1) and that the same goes for (SC_2) and WC_2.[10] But what about other relationships between these theses?

Initially, Kim (1984) and (1987) suggested that the alternative formulations of weak and strong covariance were equivalent and that so were (SC_2) and (GC). Both of these claims have come under attack in subsequent years. Van Cleve (1990) disputes the putative equivalence of (WC_1) and (WC_2), as well as Kim's argument to that effect, by objecting to Kim's use of complementation as a property-forming operation.[11] Perhaps the easiest way to motivate this objection is as follows: (WC_1), as applied to mental on physical covariance, would be trivially true if in any given possible world no two entities had the same physical properties. But intuitively, (WC_2) could still be false if in one of these possible worlds there were an entity *e* with some mental property but no physical properties (e.g., a Cartesian ego). This seems to suggest that (WC_1) does not entail (WC_2). It is easy to appreciate that, with minor complications, a similar thought experiment would militate against the purported equivalence of (SC_1) and (SC_2).

But if we take the base set to be closed under negation (as well as conjunction), *e* ends up having a physical property after all, and the conclusion that (WC_1) does not entail (WC_2) is blocked. For there is a

property in the base set that is the conjunction of the complements of all the basic physical properties, namely, $\neg B_1$ & $\neg B_2$ & . . . & $\neg B_n$, and e instantiates this "physical" property. McLaughlin's essay in this volume incorporates a comprehensive discussion of this controversy.

The purported equivalence of (SC_2) and (GC) has also come under attack, the most well-known dissent appearing in Petrie (1987).[12] Kim (1987) seemed to accept Petrie's criticisms, but this was not enough to lay the dispute to rest.[13] The discussion is continued in this volume in the essays of Klagge and McLaughlin.

Let us return to the question, Which formulation of supervenience should be preferred? pretending that it can be answered without settling questions regarding the entailment relationships between various purported formulations. In our volume, we get different answers from Bacon and Post. The former observes that weak supervenience entails strong supervenience under an understanding of properties as functions from possible worlds to individuals, but not when properties are construed as tropes. Favoring the trope account, Bacon prefers weak supervenience as the only formulation that does not entail some sort of necessary coextension. Post first defends global supervenience against charges of excessive permissiveness. Then, he contrasts it with other supervenience claims he characterizes as "individualistic" and, citing Milikan,[14] suggests that there is empirical evidence at least for some restricted global supervenience claims.

3. Supervenience and Reduction

At the same time, there is a further consideration that plays a major role in guiding various choices. As already indicated, supervenience is minimally supposed to be a nonreductive relationship, yet there is the danger that certain strong formulations of supervenience might imply reduction, which will render such formulations unsuitable for the proponents of nonreductive positions in various areas of philosophy. But first, some more background.[15]

It has now become the established lore of the supervenience literature to pinpoint Putnam (1967) and Davidson (1970) as the two main sources of the final critique of reductionism in the philosophy of mind, a critique that has had wide-ranging consequences for reductionist positions in all areas of philosophy. Both Putnam's and Davidson's arguments can be seen as undermining a necessary condition of the most popular model of reduction in the 1960s: the bridge laws of Nagel-type theoretical reductions.[16] According to the Nagel model, T_1 reduces T_2 if and only if, roughly, all the laws of T_2 (or close approximations thereof) can be derived from the laws of T_1 in conjunction with bridge laws that connect the terms of the two theories.[17] Now Putnam's

arguments can be taken to imply that such bridge laws will not be forthcoming because of the multiple realizability of the mental,[18] whereas Davidson's arguments seem to once again undermine the possibility of bridge laws by establishing the "anomaly of the mental," that is, the lack of strict lawlike relationships between the mental and the physical. On the one hand, if mental properties are multiply realizable, then no mental property will be correlated with a unique physical property;[19] on the other hand, if the mental domain is anomalous, then there will be no room for strict psychophysical correlations between mental and physical properties.[20]

Just when supervenience was becoming entrenched as the favored nonreductive notion of dependence in analytical philosophy, Kim (1983, 1984) argued that (SC_2)-type covariance between *A*s and *B*s entails the reducibility of *A*s to *B*s. The reducibility is supposed to be entailed by the fact that such a strong covariance relation entails what amounts to bridge laws or, as Kim (1990a) puts it, "strong connectibility" between theories about *A*s and *B*s, defined as follows:[21]

> T_1 is strongly connectible with respect to T_2. $=_{def}$ Each n-place predicate of T_2 has a nomological coextension in the vocabulary of T_1; that is, for each n-place predicate P of T_2, there is an n-place open sentence, P^*, of T_1 such that the following is a bridge law:
>
> $\forall x_1 \ldots \forall x_n[P(x_1 \ldots x_n) \leftrightarrow P^*(x_1 \ldots x_n)]$

The suggestion that (SC_2) implies reducibility was disputed by Teller and Post as early as 1983.[22] The key objection is that Kim's argument uses disjunction, specifically, infinitary disjunction, as a property-forming operation, whereas disjunctive properties have been held suspect for some time.[23]

In this volume, Bonevac argues that even if infinitary disjunctions are allowed, what results from strong covariance is something weaker than what is commonly understood by reduction. Strong covariance implies reduction in an infinitary language, and with such languages being incomprehensible to humans, we only get "reduction in the mind of God." Nevertheless, Bonevac suggests, this is only an *epistemological* difference: the *ontological* point, that strong covariance implies reducibility in an important sense, remains. Grimes defends the claim that similar reductive consequences also follow from global supervenience. Also arguing that neither thesis captures the intuitive idea of determination, he suggests that there are no general grounds on which to choose between (SC_2) and (GC). Macdonald attempts to give an account of psychophysical supervenience that is captured by (SC_2) without being reductive. Drawing from the example of biological properties, she argues that the existence of necessary coextensivity be-

tween two kinds of properties is not sufficient for reducibility insofar as the types of relations the supervenient properties bear to each other do not just replicate the pattern of causal relations between the physical properties that realize them.

It is worth noting here a point that is not emphasized in the supervenience literature: we have seen that if we do not assume that complementation is a property-forming operation under which all families of properties are closed, there is reason to believe that (SC_1) is not equivalent to (SC_2). If so, does (SC_1)-covariance between *A*s and *B*s entail the reducibility of *A*s to *B*s?

The answer appears to be in the negative, unless we assume the base family to be closed both under infinitary disjunction *and* complementation. To take up the second of these, for there to be a necessary coextension between *A*s and *B*s, there need to be one-way conditionals connecting *A*s to *B*s. As Kim puts it, for each property A_i in A, we need to have in each world conditionals of the form $\forall y\ (A_i y \rightarrow B^* y)$, where B^* designates the infinite disjunction of the *B*-maximal properties (i.e., the strongest consistent properties constructible in *B* relative to the property-constructing operations allowed) that ground A_i in various possible worlds. But (SC_1) can be true in models incorporating a world containing an entity with no physical properties, such as the Cartesian ego mentioned in Section 2 (again assuming that the conjunction of the complements of physical properties need not be a physical property), insofar as any two entities from the worlds in such a model are physically distinct.[24] But then, trivially, in such a world, $\forall y(A_i y \rightarrow B^* y)$ will not hold for some value of A_i, and this will be sufficient to block necessary coextensiveness of the two families of properties.

Of course, such a conclusion can be resisted by restricting the quantifiers in (SC_1) to worlds where everything has some physical properties. But as noted earlier, this opens another can of worms about the strength of the various covariance theses, specifically about the strength of the various modalities they purport to express.

4. Arguments for and against Supervenience

That a supervenience relation holds between this or that family of properties (or what have you) is a substantial thesis that goes beyond formulating various covariance principles and determining their logical properties. How does one establish such a substantive thesis? Bonevac (1988) argues that supervenience claims do not go far in supporting the ontic priority claims they are supposed to support unless backed by actual reductions. Some of the contributors in this volume

appear to disagree. Loewer argues for a thoroughgoing physicalism by defending the supervenience of any physically detectable property on basic physical properties. A crucial premise of his argument is that physics is closed and complete. This, roughly, amounts to the claim that all physical events are determined by prior physical events according to physical laws. Papineau considers an argument for a similarly strong supervenience of mental properties on the physical, which, he suggests, can be extended to show the supervenience of all real properties on the physical. Both the closed and complete character of physics, and the "manifestability of the mental" (i.e., that difference in mentally different systems will display differing physical consequences at least in some physical contexts) are key suppositions of this argument. Papineau tries to argue for the manifestability of the mental by showing that mental facts have to be realized by physical facts. Heil's contribution is a sustained effort to answer the triviality and symmetricity charges raised against physicalism in Miller (1990), where the "symmetricity charge" amounts to the claim that, given certain assumptions, physical properties also supervene on other kinds of properties.

Moser and Trout focus on global supervenience and raise problems both against the epistemic support for it and against various attempts to furnish an account of the nonreductive dependency relation it tries to capture. They conclude that such an account in terms of global supervenience that also upholds the primacy of the physical has bleak prospects. Lehrer takes issue against claims that epistemic properties supervene on nonepistemic properties and extols the virtues of the coherence theory of justification as entailing the rejection of such supervenience. The entailment follows from the fact that the notion of coherence cannot be used except in terms of epistemic concepts such as *trustworthiness* and *comparative reasonableness*. Van Cleve formulates antirealism as a thesis that claims that truth supervenes on evidence. He argues that once we focus on such a clear formulation, antirealism is faced with a number of insurmountable objections.

Ackerman and Savellos are concerned with the application of supervenience to ontological matters. Ackerman proposes a technique for transforming questions that prima facie imply the existence of certain kinds of entities (e.g., numbers, events, objects) to questions that are ontologically noncommittal and argues that the answers to the former questions supervene on the answers to the latter. Savellos examines whether events could supervene on a subvenient base consisting of objects, properties, and times, and concludes that given certain assumptions about the individual essences of events, there is no place for a nonreductive account of event supervenience.

5. Supervenience and Causation

Focusing our discussion on the philosophy of mind for the moment, if we assume that mental properties supervene on physical properties, can we also maintain our usual talk about mental causes? As we saw earlier (Loewer, Papineau), such supervenience claims are usually based on the assumption that basic physical conditions and laws are sufficient to account for any physical change that can be explained. But now, think about some mental effect that is putatively caused by a mental cause. Since the mental effect supervenes on some physical condition, and the occurrence of this physical condition can be fully explained in terms of the operation of physical laws on antecedent physical conditions, there is no work left for the original mental cause to do unless we assume large-scale causal overdetermination. Mental to physical supervenience seems to lead to epiphenomenalism about the mental.

An argument along these lines is suggested by Kim (1989a) as raising the problem of "explanatory exclusion" and is further investigated in Kim (1989b, 1990b).[25] The problem has been a source of headache for nonreductive materialists but can be appreciated by any philosopher who wishes to maintain a supervenience thesis about anything, while preserving the causal efficacy of the supervenient domain. Enç meets the challenge head-on and argues for the existence of what he calls "nonreducible supervenient causal properties." According to Enç, these properties, which include representational properties, have causal efficacy "that does not get fully accounted for by the causal role played by [their] micro base properties." Conee addresses the problem in the context of theories of meaning and retreats from supervenience to type-identity between mental and physical properties. He argues that the standard multiple-realizability arguments, contrary to common belief, do not present type materialism with insurmountable problems.

6. Is Supervenience Enough?

On a final note, we would like to consider two ways in which supervenience has been criticized for not delivering what it promises. To begin with, it has been suggested that without the specification of the particular determination or dependence relationship that underlies it, the claim of supervenience is "simply an empty sound expressing a faith that two levels of properties are somehow related."[26] As Schiffer puts it, "Invoking a primitive metaphysical relation of supervenience to explain how non-natural moral properties were related to physical

properties was just to add mystery to mystery, to cover one obscurantist move with another."[27]

Let us, for the sake of simplicity, focus on global covariance between the mental and the physical. The problem that the preceding complaints raise seems to be the following: establishing that worlds alike in physical respects have to be alike in mental respects does not seem, by itself, to explain either why this relationship obtains between the mental and the physical or why the physical is in some sense prior to the mental. The proponent of the covariance just presents such a relationship as a brute metaphysical fact and often accompanies his claim with the somewhat dogmatic pronouncement that the physical is, somehow, ontologically the more basic of the two. But a materialist who wants to make assertions about the ontic priority of the physical has to say more than this to explain *how* and *why* the covariance relation between the mental and the physical amounts to dependency. Perhaps he can appeal to microdetermination, conceptual dependence or part–whole relationships to make his point, but one must appeal to something more than mere covariance.[28] The point can be generalized, mutatis mutandis, to covariance between other domains.

One might attempt to minimize the force of this objection by granting that it would be desirable to explain why and how supervenience obtains by appealing to more familiar dependence or determination relationships *when this is possible*. At the same time one can insist that sometimes supervenience can be a brute fact as little in need of explanation as the basic laws of physics or logic.[29] That is, if we cannot discover some familiar dependence or determination relationship between *A*s and *B*s, but we are still confronted with the fact that *A*s covary with *B*s (and not vice versa?), we might after all be forced and entitled to postulate an unexplained, perhaps primitive dependence relationship between them. Needless to say, the final word on whether this is a fully satisfactory maneuver is not in.

At this point, we can turn to an alternative criticism of supervenience. It has been frequently observed that satisfaction of even the strongest covariance thesis considered in Section 2 is by no means sufficient to establish dependence or determination between the properties that covary in this manner.[30] For example, (SC_2)-type mental to physical covariation can be satisfied in a single world model in which psychophysical parallelism is true. Is this not, then, a serious shortcoming of the theses under consideration? And more importantly, if all supervenience claims are based on the existence of such covariations that need not be explained by appeal to a dependency, does this not undermine the recently made suggestion that such covariation entitles us to postulate an unexplained, perhaps primitive dependence relationship between the covarying domains?