

Supervenience, Emergence, and Reduction¹

ANSGAR BECKERMANN

At first glance, the theories of emergence and supervenience as mind-body theories seem to be very close relatives. Both try to avoid the extremes of outright dualism with its assumption that life and mind can only be explained by non-material entities such as entelechies or spiritual souls as well as that of “totalistic” physicalism – or mechanism, as the emergentists used to say – with its adverse assumption that all entities in the world are definable in terms of physical entities or are at least reducible to them. That is, the theories of emergence and supervenience are both versions of what is now usually called “physicalism without reductionism”. But, as I shall try to show in this paper, emergence and supervenience are by no means identical twins. There are remarkable differences that are well worth examining. These differences will become clear when we look more closely at the theories of supervenience and emergence themselves as well as the notions of reduction that are associated with them.

1. *Supervenience*²

The idea of supervenience seems to have appeared for the first time in the writings of the moral philosophers G. E. Moore and R. M. Hare. Both authors shared the idea that although moral concepts or properties such as being a good person or being a right act are not definable in terms of

¹ I am very grateful to Jaegwon Kim and Brian McLaughlin for their helpful comments on an earlier draft of this paper. Especially the very detailed remarks by Brian McLaughlin have helped me a lot to understand a variety of issues much better than before. Quite a number of notes are due to his elucidating criticism (compare e. g., notes 7, 8, 10, 11, 18, and 19). In addition I would like to thank Antonia Barke and David J. Drinkwater for proof reading and correcting my English.

² This section in which I rely heavily on the works of J. Kim listed in the bibliography is meant to expound the main ideas connected with the notion of supervenience. Therefore, it will mainly contain material which most of the readers will be well acquainted with.

natural properties, these properties nonetheless strongly depend on natural properties in the following way: if any two persons or acts are exactly alike with respect to all their natural properties they must also be alike with respect to their moral properties. If all natural properties of an object are fixed its moral properties are fixed, too. There can be no moral difference between two objects or situations without there also being an underlying natural or physical difference. But this does not imply that moral properties are definable in terms of, or reducible to, physical properties.

D. Davidson probably was the first to use this idea of "dependency without reduction" to characterize the relation between mental and physical phenomena. That is, he was the first who used the term "supervenience" to characterize a relation of this kind.

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 respect, or that an object cannot alter in some mental respect without altering in some physical respect. (1970, p. 214)

It is obvious, that Davidson here uses the term "supervenience" to refer to a dependency relation which may hold between two families of properties even if there are no laws connecting these properties and which, therefore, *a fortiori* does not imply that the properties of the one family are reducible to the properties of the other.

But what would such a dependency relation look like? How are we to understand the two clauses "that there cannot be two events alike in all physical respects but differing in some mental respect" and "that an object cannot alter in some mental respect without altering in some physical respect"? If we ignore the difference between events (mentioned in the first clause) and objects (mentioned in the second clause) and concentrate on objects, the first clause simply seems to imply that mental properties supervene on physical properties in that any two objects which share all physical properties also share all mental properties or, to put it the other way round, in that there are no two objects which share all their physical properties, but differ in some mental property. Supervenience according to the first clause thus amounts to what has been called "weak supervenience". This concept has been made more precise by J. Kim in the following way:

(WS) Let A and B be two sets of properties: then A weakly supervenes on B if and only if necessarily any two things that have the same properties in B have the same properties in A (that is, B-indiscernibility entails A-indiscernibility).

If we assume that the set B is closed under the usual Boolean operations complementation, conjunction, and disjunction (and perhaps some other operations such as infinite conjunction and disjunction), weak supervenience can be equivalently defined as follows:

(WS') A weakly supervenes on B if and only if 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 , y also has F .

But weak supervenience, defined in either of these two ways, can be shown to be a rather weak relation, probably much too weak to catch the intuitive idea of determination or dependence that most philosophers have in mind when they use the term "supervenience". For if supervenience really was a dependency relation it should support counterfactual claims.

If I feel a pain here and now, weak supervenience, however, only implies that there must be a physical condition P such that it is true that I am now in P and that the generalization "Everyone who is in P feels pain" *happens* to be true in the actual world. But even in the nearest possible world the physical conditions underlying pain could be quite different. That is, according to weak supervenience, it could well be that, even in the nearest possible world, I feel no pain though being in P . And if this were the case, the subjunctive counterfactual "If I were in P I would feel pain" would be actually false.

Kim stresses much the same point, when he points out that weak supervenience is compatible with each of the following three claims:

- (a) There is another possible world, in which there exists an exact physiological replica of me which does not feel any pain.
- (b) There is another possible world, in which there are creatures who are exactly like us in all physiological respects, but who do not have any mental properties at all.
- (c) There is another possible world in which not only entities with a certain physiological makeup feel pains, but also trees and rocks and lakes.

Obviously, all this is due to the fact that the properties G_i may differ radically from world to world or, to put it in other words, that each of the generalizations "If x has G_i , x also has F " may only be accidentally true in the corresponding possible world. That is, that all these generalizations may lack modal force. If we want to strengthen the concept of supervenience in such a way that a real dependence of supervenient on subvenient properties results, it is, therefore, natural to add exactly this modal force and define supervenience in a way which Kim calls "strong supervenience":

- (SS) A strongly supervenes on B if and only if 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 , y has F .

But if weak supervenience was too weak to catch the intuitive idea of determination or dependence associated with the idea of supervenience, strong supervenience seems to be too strong, since at least one can argue — as Kim did in a series of papers — that strong supervenience is incompatible with antireductionism. To see how Kim's argument works we have to be a little more technical. Let us assume that the set of subvenient properties B comprises the three basic properties G_1 , G_2 and G_3 and any properties constructible from these by the mentioned operations of complementation, conjunction, and so on. Further let us assume that a is an object of the domain under consideration which has the properties G_1 and G_2 , but lacks the property G_3 . If we now define B_1 as the conjunction of the B-properties a has and of the complements of the B-properties a lacks, that is, if we define B_1 by " x has B_1 iff x has G_1 and G_2 and $\neg G_3$ ", then a obviously has B_1 .³ But if a also has F and F is a member of a family A of properties that strongly supervene on B, then the generalization

- (1) For all x : if x has B_1 , then x has F

must hold with necessity. For B_1 is a B-maximal property in the sense that B_1 determines which B-properties an object has and which it lacks. So, if a has B_1 and b has B_1 , a and b are B-indiscernible and must therefore have the same A-properties. But B_1 is of course not the only B-maximal property. " x has G_1 and $\neg G_2$ and $\neg G_3$ ", e. g., refers to another property of this kind. It is easy to see, that there are exactly eight such B-maximal properties B_1, \dots, B_8 , and that for each of these properties there exists a law like (1) which tells us whether an object that has one of these eight properties, also has or lacks the A-property F . If we assume that objects which have one of the properties B_1, B_3 , or B_8 also have the property F and that all other objects lack this property and if we define the property B_F by " x has B_F iff x has B_1, B_3 or B_8 " and the property $B_{\neg F}$ by " x has $B_{\neg F}$ iff x has B_2, B_4, B_5, B_6 , or B_7 ", we, therefore, have the following two nomological biconditionals:

- (2) For all x : x has B_F iff x has F .
 (3) For all x : x has $B_{\neg F}$ iff x has $\neg F$.

³ If B is not finite as in our example, but infinite, B_a of course will be an infinite conjunction.

And this at least seems to imply that both F and $\neg F$ are reducible to certain B-properties.

But this conclusion might be too hasty, since one can argue (a) that nomological biconditionals are neither necessary nor sufficient for reduction (I will come back to this point in the section on reduction below) or (b) that, even if one grants that they are, such biconditionals imply that the (mental) property F is reducible to the (physical) property G only in case that the predicate " G " is not only definable in physical terms, but is itself a natural kind predicate in physics.⁴ That G determines a natural kind or that " G " is a natural kind predicate in this context is understood to mean that there are laws of physics such that objects or events fall under these laws by virtue of instantiating this property or by virtue of satisfying this predicate. According to this account the biconditionals (2) and (3) probably do not imply that F and $\neg F$ are reducible to certain B-properties, since there are lots of reasons for doubting that the predicates B_F and $B_{\neg F}$ are natural kind predicates in this sense. So, strong supervenience may after all still be compatible with antireductionism.

Independent of these arguments some authors⁵ felt that the notions of weak and strong supervenience do not catch the intuitive idea behind the concept of supervenience in general, since they concentrate too much on particular objects and events and on the question how the properties of such objects or events are related to each other. These authors, therefore, developed an alternative account which starts instead from the comparison of whole worlds and which, for this reason, has been called "global supervenience". According to this account, the supervenience of the mental on the physical should rather be explained by saying that *worlds* that are physically indiscernible are also psychologically indiscernible. And even the supervenience of the moral on the nonmoral should be explained in much the same way: There could not be two *worlds* which are indistinguishable in every nonmoral feature and yet differ in some moral respect. Thus, the concept of global supervenience can be defined like this

(GS) A globally supervenes on B if and only if any two worlds w_1 and w_2 which are B-indiscernible are also A-indiscernible.

What speaks in favour of the so defined concept of global supervenience? In the first place certainly the fact that the mental states of a person in many cases seem to depend not only on what is going on in her head or her CNS, but also on her history or on the environment she lives in. Let us grant for the moment that to have the belief that p is to stand in a

⁴ Fodor (1974), Teller (1984).

⁵ Cf. e. g. Haugeland (1982), Horgan (1982).

certain functional relation to a corresponding mental representation which means that p . Now, mental representations are physically realized symbols, i. e., they presumably are certain brain states. But what these brain states mean depends heavily on the history of a person, under what conditions she acquired this brain state for the first time and how this brain state is now causally related to certain features of her environment.⁶ It is, therefore, quite imaginable that two persons A and B stand in the same functional relation to the same brain state and yet have different beliefs, since these brain states have different meanings. If we want to make sure, that A and B have the same beliefs (and desires, etc.) we must, therefore, not only assume that they are in the same brain states, but also that they had the same history and that they have been living in the same environment. And this is certainly guaranteed if not only A and B , but also the whole worlds these two persons live in are physically indiscernible.⁷

Apart from these considerations there is also a more theoretical reason for thinking that global supervenience is preferable to weak and especially to strong supervenience, since it at least seems to be the more convincing explication of what could be meant by "determination *without* reduction". For if, e. g., mental properties globally supervene on physical properties, the distribution of the physical properties in a way determines the distribution of the mental properties, since global supervenience implies that if all physical properties of all objects in a world are fixed, the mental properties of these objects are fixed, too. But it does not imply that there are any property-to-property connections whatsoever, since it is entirely compatible with there being two particular objects a and b which share all physical properties, but differ in at least one mental respect.

On the other hand, however, it is precisely this fact that reveals the great weakness of global supervenience. For it makes clear that the determination that is implied by global supervenience is a kind of determination which hardly deserves this name. The point is, as J. Kim has put it, that

it is consistent with this version of materialism for there to be a world which differs physically from this world in some most trifling respect (say, Saturn's rings in that world contain one more ammonia molecule) but which is entirely devoid of consciousness, or has a radically different, perhaps totally irregular, distribution of mental characteristics over its inhabitants (say, creatures with brains have no mentality while rocks are conscious). (1987, p. 321)

⁶ This, I think, has been convincingly shown by Putnam's twin earth cases and Burge's arguments against Psychological Individualism. Cf. Putnam (1975) and Burge (1979, 1986).

⁷ Instead of going global, however, one could also stay with strong supervenience and try to solve the mentioned problems by admitting relational properties.

If only a world differs from this one in some physical respect, however minuscule or seemingly irrelevant, the mental properties of the objects in this world could differ from the mental properties of the objects in our world as radically as you please.

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. (ibid.)

So, if we really want supervenience to be a dependency relation, strong supervenience seems to be the only viable option left.⁸

2. *Emergence*

When I now turn to the second of the three key notions of this paper, the notion of emergence, I shall concentrate on the theory of C. D. Broad since the notion of emergence which Broad developed in his book *The Mind and Its Place in Nature* is very elaborate and most interesting. Broad set forth his theory of emergence in the context of a discussion on what he called "Substantial Vitalism", a theory which claimed that the presence of a particular entity, often called an "Entelechy", is a necessary factor in explaining the characteristic behavior of living bodies, an entity that does not occur in inorganic matter or in bodies which are no longer alive.

In order to be able to show what is wrong with "Substantial Vitalism" Broad gives a taxonomy of the kinds of theories that can be advanced if one is concerned with the explanation of the characteristic behavior of certain kinds of things or systems. Principally, there are two opposite types of view.

According to the first type of view the characteristically different ways of behaving are regarded "as absolutely unanalysable facts which do not depend *in any way* on differences of structure or components". Obviously, this type of view is absurd with regard to the behavior of living beings. For we all know that this behavior is at least in part dependent on the physiological structure of these beings. If this structure is impaired or destroyed the behavior will no longer occur in the same way. Moreover, this first type of view

can hardly be counted as one way of *explaining* differences of behaviour, since it consists in holding that there are certain differences which cannot be explained, even in part, but must simply be swallowed whole with that philosophical jam which Professor Alexander calls 'natural piety'. (p. 55)

⁸ Even this, however, is no longer undisputed. Cf. Grimes (1988), Kim (1990) and the short remarks in my introduction to this volume pp. 13–14.

The second type of view comprises all theories which — just on the contrary to the first type of view — assume that the characteristic behavior of certain objects can, at least in part, be explained in terms of their structure or components or both. These theories can be further divided into three classes: component theories, emergent theories, and mechanistic theories.

The characteristic feature of component theories is the claim that the behavior of a certain class of objects is, in part, to be explained by “the presence of a peculiar *component* which does not occur in anything that does not behave in this way” (p. 55). Theories of this kind are by no means *per se* unscientific. So, if “Substantial Vitalism” is often regarded as the *paradigm of an unscientific theory*, this is not due to the fact that “Substantial Vitalism” is a component theory, but to the fact that the particular components it postulates are for at least three reasons rather odd entities.

(i) Neither entelechies nor other entities of a similar kind have ever been isolated. They are “*purely hypothetical*” entities “in a sense in which an as yet unisolated but suspected chemical element is not” (p. 57). (ii) It is certainly true, not of chemical elements, but of a number of very important groups of elements that they cannot exist in isolation. But these groups unlike entelechies are not mysterious simple entities; they are groups of known elements; and their inability to exist by themselves can itself be explained by the theoretical principles of chemistry. (iii) Entelechies are supposed to be entities of an entirely different kind than normal material entities. It is not even clear whether they are literally in space at all. But if they are not, it is not easy to understand what is meant by saying that they are *parts of living bodies*, that a living body is a compound of an entelechy and a material structure. This last point perhaps is the most convincing. But this paper is not meant to be a discussion of vitalism. So we should go on to the other types of theories Broad discusses.

The rivals of component theories, emergent theories and mechanistic theories, concur in denying “that there need be any peculiar *component* which is present in all things that behave in a certain way and is absent from all things which do not behave in this way. [They both say] that the components may be exactly alike in both cases, and [try] to explain the difference of behaviour wholly in terms of difference of structure.” (pp. 58 f.) But although they share this very general point of view they differ very radically in their view of the laws which connect the properties of the components with the characteristic behavior of the complex wholes which these components make up. Emergent theories claim that

the characteristic behaviour of the whole *could* not, even in theory, be deduced from the most complete knowledge of the behaviour of its components, taken

separately or in other combinations, and of their proportions and arrangements in this whole. (p. 59)

Broad hesitates to give examples of cases that could be counted as emergent in this sense since it is, as he admits, a controversial matter whether there actually are such cases. But it is obvious from his writings that Broad himself was convinced that the characteristic behavior of almost all chemical compounds is emergent in the sense explained. At least he remarks "that, so far as we know at present, the characteristic behaviour of Common Salt cannot be deduced from the most complete knowledge of the properties of Sodium in isolation; or of Chlorine in isolation; or of other compounds of Sodium, such as Sodium Sulphate, or of other compounds of Chlorine, such as Silver Chloride" (p. 59).

In contrast to emergent theories mechanistic theories claim that

the characteristic behaviour of the whole is not only completely *determined* by the nature and arrangement of its components; in addition to this it is held that the behaviour of the whole could, in theory at least, be *deduced* from a sufficient knowledge of how the components behave in isolation or in other wholes of simpler kind. (p. 59)

Artificial machines are, according to Broad, the paradigmatic cases of things which admit an entirely mechanistic explanation. His own example is that of a clock which behaves in a characteristic way, i. e., its hand moves around steadily in such a way that every full hour it stands exactly above the corresponding mark on its face. Obviously, there is no reason to suppose that this behavior is due to a peculiar entity that is not present in anything but clocks. But what is more, the behavior of the hand can be deduced completely from the particular arrangement of springs, wheels, pendulum, etc., and from the general laws of mechanics and physics which apply to all kinds of material objects and not only to clocks or their components. So, the behavior of a clock is not emergent in Broad's sense.

What is essential for emergent and mechanistic theories is summarized by Broad as follows:

Put in abstract terms the emergent theory asserts that there are certain wholes, composed (say) of constituents A, B, and C in a relation R to each other; that all wholes composed of constituents of the same kind as A, B, and C in relations of the same kind as R have certain characteristic properties; that A, B, and C are capable of occurring in other kinds of complex where the relation is not of the same kind as R; and that the characteristic properties of the whole R(A, B, C) cannot, even in theory, be deduced from the most complete knowledge of the properties of A, B, and C in isolation or in other wholes which are not of the form R(A, B, C). The mechanistic theory rejects the last clause of this assertion. (p. 61)

The concept of emergence, or to be more precise: the concept of an emergent property that Broad has in mind therefore seems to be this:

- (E) Let S be a system made up of the constituents C_1, \dots, C_n standing in a certain relation R^9 to each other (let us call $[C_1, \dots, C_n; R]$ the microstructure of S); then F is an emergent property of S iff (a) there is a law to the effect that all systems with this microstructure have F , but (b) F cannot, even in theory, be deduced from the most complete¹⁰ knowledge of the properties of the components C_1, \dots, C_n in isolation or in systems with a different microstructure.

If this is a correct explication of what Broad had in mind, the characteristic feature of emergent properties is not that there can be no laws connecting these properties with the having of a certain microstructure. On the contrary. According to Broad for every emergent property F there must be at least one corresponding microstructure such that it is a law of nature that all objects that have this microstructure also have property F . In other words, according to Broad, emergent properties must strongly supervene on microstructural properties.¹¹ For otherwise the presence of such properties could in no way be explained by reference to the corresponding microstructures. Strong supervenience, however, does not preclude the possibility that emergent properties “cannot, even in theory, be deduced from the most complete knowledge of the properties of the components C_1, \dots, C_n in isolation or in systems with a different microstructure”. And it is exactly this feature that is decisive for a property to be emergent.

But why does Broad use this complicated way of putting things? Why does he not simply say that a property is emergent if it cannot, even in theory, be deduced from the most complete knowledge of the properties of the components C_1, \dots, C_n , leaving out the additional clause “in isolation or in systems with a different microstructure”? To try and give an exhaustive answer to this question would by far exceed the scope of this paper. But I shall try at least to hint at the reasons which led Broad to a formulation of this kind. I think that the main point he had in mind was a point very clearly stated by C. G. Hempel and P. Oppenheim about 20 years later in their article “Studies in the Logic of Explanation” (1948).

⁹ This relation, I think, should be defined in purely spatial terms, since otherwise one could put too much into it. (Cf. the remarks on “ad hoc properties” and “ad hoc laws” below.) But I don’t want to elaborate this point further here.

¹⁰ For Broad it is selfunderstood that the most complete knowledge of the properties of the components of a system also comprises knowledge of the laws which hold for entities that have these properties. Cf. below pp. 105–106.

¹¹ Here and elsewhere I am presupposing that the second “necessarily” in the definition (SS) of strong supervenience expresses physical, not logical necessity.

In their remarks on the concept of emergence Hempel and Oppenheim write with reference to a corresponding remark by Grelling:

If a characteristic of a whole is counted as emergent simply if its occurrence cannot be inferred from a knowledge of all the properties of its parts, then, as Grelling has pointed out, no whole can have any emergent characteristics. Thus ... the properties of hydrogen include that of forming, if suitably combined with oxygen, a compound which is liquid, transparent, etc. Hence the liquidity, transparency, etc. of water *can* be inferred from certain properties of its chemical constituents. (p. 260)

If the concept of emergence is not to be vacuous, such inferences must be blocked in one way or other. Broad's formula, I think, is meant to do just this. What he wants to make sure of, is that the properties of the components of a system which we can rely on in trying to deduce one of its macroproperties do not comprise properties of the kind mentioned by Hempel and Oppenheim. But obviously he is overstating his point, since it might well be that we come to know some of the *basic* properties of the components of a system only if we observe these components in certain hitherto unencountered relations. And it certainly would be unfair to exclude these properties as well. I mention only the weak and strong interaction between certain elementary particles as examples of what I have in mind. It, therefore, would be more appropriate to leave out the epistemological aspect implicit in Broad's formula and concentrate on what is decisive: that the properties of the components from which one could try to deduce a certain macroproperty F should not comprise "ad hoc properties" like the one that components of this kind standing in the relation R to each other make up a system which has exactly this property. But what other formulation could do the job? I propose simply to use the term "basic" to solve this problem knowing that this is not a very satisfactory solution. For such a solution would presuppose an adequate account of what basic properties are. But I don't even want to sketch such an account, since perhaps the best we can expect in this context is a list of the properties that are admissible and in this sense basic.

If we accept this proposal our first definition of the concept of an emergent property comes to be like this:

(E') Let S be a system having the microstructure $[C_1, \dots, C_n; R]$, then F is an emergent property of S iff (a) there is a law to the effect that all systems with this microstructure have F , but (b) F cannot, even in theory, be deduced from the most complete knowledge of the *basic* properties of the components C_1, \dots, C_n .

But even this formulation is not yet entirely satisfactory since the concept of emergence can be made vacuous not only by using "ad hoc properties",

but also by using "ad hoc laws". For just as we can infer water to have the properties of being liquid and transparent if we may rely on the property hydrogen has "of forming, if suitably combined with oxygen, a compound which is liquid, transparent, etc." we can infer this property if — relying only on the basic properties of hydrogen and oxygen — we are allowed to use a law like "If hydrogen and oxygen are combined in manner *R* they form a compound which is liquid, transparent, etc.". So, the possibility of using such laws must also be precluded. At least, if these laws are so to speak brute laws that cannot be explained any further.

Broad himself stresses the fact that we must rely on laws when we want to deduce the properties of certain wholes, since "in *no* case could [the properties] of a whole composed of certain constituents be predicted *merely* from a knowledge of the properties of these constituents, taken separately, and of their proportions and arrangements in the particular complex under consideration" (p. 63). If this sometimes seems to be possible this is only because we use hidden premises which we are so familiar with that we do not notice them any more. These premises stem from the examination of "other complexes in the past". We found a *general law* connecting the properties of these wholes with the properties of their constituents, and we assume that this law will also hold for the new complex that is now the object of our investigations.

For purely dynamical transactions this assumption is pretty well justified, because we have found a simple law of composition and have verified it very fully for wholes of very different composition, complexity, and internal structure. It is therefore not particularly rash to expect to predict the dynamical behaviour of any material complex under the action of any set of forces, however much it may differ in the details of its structure and parts from those complexes for which the assumed law of composition has actually been verified. (pp. 63 f.)

As to chemical compounds on the other hand we have — according to Broad — *no right to assume that their properties are deducible from general laws applying to all different sorts of compounds which have in a way the same or similar constituents. On the contrary, if we want to find what properties a compound has, there is no other way than to study samples of that particular compound.*

The essential point is that it would ... be useless to study chemical compounds in general and to compare their properties with those of their elements in the hope of discovering a *general law* of composition by which the properties of *any* chemical compound could be foretold when the properties of its separate elements were known. So far as we know, there is no general law of this kind. (p. 64)

This does not mean that the properties of silver-chloride are not "determined" by those of silver and chlorine; it only means that the law connecting "the properties of silver-chloride with those of silver and of

chlorine and with the structure of the compound is, so far as we know, an *unique and ultimate law*"¹² (p. 65).

By saying that this law is "unique and ultimate" Broad means that (a) it "is not a special case which arises through substituting certain determinate values for determinable variables in a general law", that it (b) "is not a special case which arises by combining two more general laws" and that it is (c) "a law which could have been discovered *only* by studying samples of silver-chloride itself, and which can be extended inductively *only* to other samples of the same substance." (p. 65)

If we contrast the case of the properties of silver-chloride with the case of the properties of a clock the difference becomes even clearer. For these properties can be predicted even by a man who has never seen a clock in his life.

Provided he is told how it is constructed, and that he has learnt from the study of other material systems the general rules about motion and about the mechanical properties of springs and of rigid bodies, he can foretell exactly how a system constructed like a clock must behave. (p. 65)

I think the lesson from all this is clear. The laws we may rely on when we try to deduce the properties of a whole from the properties of its parts must not be "unique and ultimate" laws. Otherwise there would not be any emergent properties at all. For, according to Broad, it is part of the concept of an emergent property that there is a law to the effect that all systems with a certain microstructure have this property. Therefore, only general laws are admissible which do not apply only to those kinds of wholes which are the actual objects of investigation. Taking this into account we finally arrive at the following definition of the concept of an emergent property.

(E") Let S be a system having the microstructure $[C_1, \dots, C_n; R]$, then F is an emergent property of S iff (a) there is a law to the effect that all systems with this microstructure have F , but (b) F cannot, even in theory, be deduced from the basic properties of the components C_1, \dots, C_n and a general theory of components of this kind which contains no unique and ultimate laws which apply only to systems which have the same microstructure as S .

Now, a question remains. What here is meant by "deducing a property"? Since the answer to this question is strongly connected to the problem of reduction, I will address it in the course of the next section.

¹² Of course, the further development of chemistry has proven Broad to be wrong on this at least for some properties of silver-chloride.

3. Reduction

In the foregoing remarks on supervenience and emergence, the problem of reduction and reductionism was touched on only in an unsystematic manner. It is time to address this problem in a more systematic way. According to the classic Nagelian account, reduction is a relation which holds between two theories T_1 and T_2 if the laws of theory T_1 can directly or with the aid of so called "bridge laws" be deduced from the laws of theory T_2 . Bridge laws will be necessary in most cases since the vocabulary of T_1 will in general be different from the vocabulary of T_2 so that for that reason alone a direct deduction will not be possible. Bridge laws are usually thought to have the form of nomological biconditionals which connect one term of T_1 with a corresponding term of T_2 or of equations which do the same job for a whole family of terms. The paradigm case is the reduction of thermodynamics to statistical mechanics.

If thermodynamics is to be reduced to statistical mechanics its laws, such as Boyle-Charles's law

$$(1) \quad p * V = N * k * T,^{13}$$

must be deducible from the laws of statistical mechanics. But obviously this cannot be done in a straightforward way, since neither " p " ("pressure") nor " T " ("temperature") belong to the vocabulary of the latter science. From the laws of statistical mechanics, however, the following law can be deduced:

$$(2) \quad M * V = 2/3 * N * E,$$

where M is the average of the momenta transferred from the molecules of the gas to the walls of the container and E the mean kinetic energy of these molecules. From this law we can deduce Boyle-Charles's law with the aid of the following two bridge laws:

$$(3) \quad p = M$$

$$(4) \quad 2/3 * E = k * T.$$

These equations give us the connection needed between the terms " p " and " V " of thermodynamics and " M " and " E " of statistical mechanics. So, the essential role which bridge laws play in the reduction of the former theory to the latter seems to be obvious.

¹³ Here N is the number of the molecules of the ideal gas under consideration in the volume V , k is the Boltzmann constant, and T is the absolute temperature of the gas.

It is, I think, a remarkable fact that in the context of the debate on supervenience, reduction or reducibility is almost entirely identified with the existence of bridge laws, that is, of nomological biconditionals which yield the necessary coextensivity of two terms that belong to the vocabulary of different theories or to different families of predicates. But even in the classic Nagelian account, the pure *existence* of bridge laws is not enough to establish reduction. These bridge laws must do some work, that is, they must enable us to deduce the laws of one theory from the laws of another. This ingredient of the notion of reduction has perhaps been lost sight of in the debate on supervenience, since supervenience is only concerned with the relations between families of properties or predicates and not with the relations between different theories.

This sole concentration on bridge laws is, however, very misleading, since the discussion over the last decades has shown that bridge laws do not play such an essential role in reduction as the classic account supposed. One of the results of this discussion was that it is not essential for the reduction of one theory T_1 to another T_2 that the laws of T_1 themselves can be deduced from the laws of T_2 . What really is essential is that one can deduce from T_2 something that is commonly called an *image* of T_1 .¹⁴ In other words, it is essential to show that there are properties which we can refer to by terms definable in terms of T_2 which play (almost) the same role as the properties which are referred to by the terms of T_1 . And it is just this that is shown by the possibility of deducing an image of T_1 from T_2 .

To show that thermodynamics is reducible to statistical mechanics, therefore, is to show that there are properties expressible in terms of statistical mechanics — e. g., the properties referred to by " M " and " $2/3 k * E$ " — that play exactly (or at least approximately) the same role as the properties referred to by the thermodynamical terms " p " and " T ", that is, to show that the former properties stand in the same relations to each other and to the corresponding observable effects as the latter ones. This, obviously, can only be done by deducing an image of thermodynamics from statistical mechanics where such an image is defined as the set of sentences that results if we replace in each law of thermodynamics the terms " p " and " T " by the terms " M " and " $2/3 k * E$ ". In order to reduce

¹⁴ This view — which has also been adapted by P. M. Churchland and P. S. Churchland — has been especially elaborated by C. A. Hooker in (1981). Its departure from the original Nagelian account has mainly been due to the realization that in most cases (e. g., the reduction of Kepler's laws to Newtonian mechanics) it is not possible to deduce exactly the laws of T_1 from T_2 , but only laws to which the laws of T_1 can be shown to be close approximations.

thermodynamics to statistical mechanics one, therefore, has to deduce from the latter theory not only the law (2) which is the image of Boyle-Charles's law, but also images of all other laws of thermodynamics. That is, one also has to deduce laws like

- (5) If a is an ideal gas contained in a vessel and the mean kinetic energy of the molecules of a is $2.07 * 10^{-23} * x$ Joule and if b is a mercury thermometer which is in thermodynamical equilibrium with a , then the mercury column of b will match the mark of x degrees Kelvin.

For this is the image of the thermodynamical law

- (6) If a is an ideal gas contained in a vessel and the temperature of a is x° K and if b is a mercury thermometer which is in thermodynamical equilibrium with a , then the mercury column of b will match the mark of x degrees Kelvin.

All these deductions, however, can be done without the aid of any bridge laws at all. So, bridge laws play no essential role in reduction.

This, I think, is the right moment to come back to the question of what it could mean to deduce the property of a system from its microstructure, i. e., from its components and the way these components are related. For there is, as the following will show, a close connection between the answer to this question and the foregoing remarks on reduction in general. In recent years R. Cummins has stressed the point that there is something like the deduction or, as he calls it, explanation of properties. And it was also Cummins who made it clear that these explanations of properties have to be carefully distinguished from the explanations of events or, as he says, changes.

Many scientific theories are designed to explain change. The point of what I call a *transition theory* is to explain changes of state in a system as effects of previous causes ... The emphasis is on what will happen *when* (...). (1983, p. 1)

But the explanation of changes is not the exclusive aim of science, probably it is not even its most prominent aim. Many of the most valuable scientific theories — such as the theory of elements in chemistry or Rutherford's and Bohr's theories about the structure of atoms — are better understood as aiming at the explanation of the properties of, e. g., atoms or compounds of atoms.

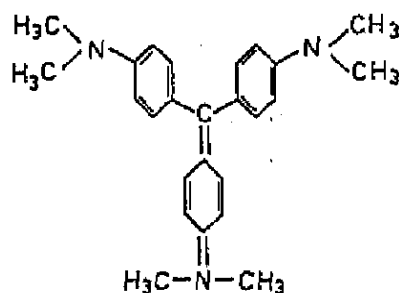
Many scientific theories are not designed to explain changes but are rather designed to explain properties. The point of what I call a *property theory* is to explain the properties of a system not in the sense in which this means 'Why did S acquire

P?' or 'What caused S to acquire P?' but, rather, 'What is it for S to instantiate P?', or, 'In virtue of what does S have P?' ... Many of the most pressing and puzzling scientific questions are questions about properties, not about changes. We know a lot about what causes pain, but there is no very good theory of how pain is instantiated. Good property theories are wonderfully satisfying: we know how temperature is instantiated, how inheritance is instantiated, how electricity is instantiated, how solubility is instantiated. (1983, pp. 14 f.)

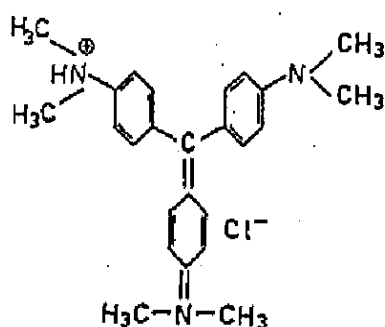
Transition theories and property theories are thus best understood as answering different types of questions: questions like "What caused x to have F ?" on the one hand, and questions like "In virtue of what does x have F ?" on the other. Let us take as an example the color and change of color of crystal violet.¹⁵ Crystal violet is a chemical compound whose structure is characterized by the formula given in Fig. 1 a. A solution of crystal violet usually has a blue-violet colour which can be put down to the fact that crystal violet absorbs the yellow-orange part of the spectrum of visible light. This in turn is due to the fact that because of the energy of light from this part of the spectrum striking the crystal violet molecules, certain electrons are raised to a higher energy state. In order for this to happen it is necessary that the three nitrogen atoms present in each crystal violet molecule are able to acquire positive charges. Now, if a little hydrochloric acid is added to the crystal violet solution one nitrogen atom in each crystal violet molecule takes up a proton, thereby preventing this atom from acquiring a positive charge (Fig. 1 b). Because of this the colour changes to green. Upon further addition of hydrochloric acid, all three nitrogen atoms take up a proton and the colour disappears completely (Fig. 1 c).

With regard to this example both kinds of questions can be asked: "Why did the color of the liquid change from violet to green?" or "What was the cause of its becoming green?", and "In virtue of what does the liquid has the color violet at one time and the color green at another?". The answer to the first kind of question must consist in stating the event which caused the change of color, that is, the addition of hydrochloric acid. Answers to questions of the second kind on the other hand will in most cases — as in the case of our example — take the form of *microreduction*. That is, they will be attempts to deduce the property of an object from the properties of its parts and their manner of organisation, very much in the sense of C. D. Broad. According to Cummins, such answers will be successful

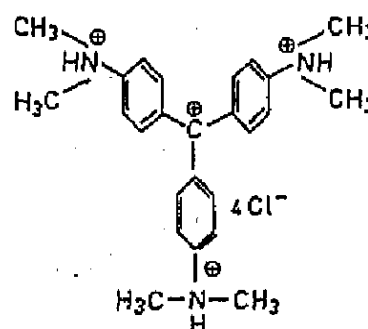
¹⁵ Cf. Glöckner (1973, pp. 136 ff.).



a. all three nitrogen atoms are able to acquire positive charges (color: blue-violet)



b. only two nitrogen atoms are able to acquire positive charges (color: green)



c. no nitrogen atom is able to acquire a positive charge (colorless)

Figure 1¹⁶

when we come to see that something having the kinds of components specified, organized in the way specified, is bound to have the target property. (p. 17)

But what is meant here by saying that systems which consist of certain components organized in a certain way "are bound" to have a certain property? Cummins seems to believe that we know that an object having a certain microstructure is bound to have property F if we know that a corresponding law is true, that is, a law like

(7) If x consists of the components C_1, \dots, C_n organized in manner R , i.e., if x has microstructure $[C_1, \dots, C_n; R]$, then x has property F .

For, according to Cummins, property explanations can be represented in accordance with the D-N-model of Hempel and Oppenheim:¹⁷

¹⁶ Cf. Glöckner (1973, p. 141).

¹⁷ Although he claims that this kind of representation may be misleading since it obscures the differences between property explanations and transition explanations.

- (8) (i) If x has microstructure $[C_1, \dots, C_n; R]$, then x has property F .
 (ii) S has microstructure $[C_1, \dots, C_n; R]$
-
- (iii) S has property F .

But this, I think, cannot be the whole truth since what we want to know if we ask for the explanation of a property is not *that* a law such as (7) is true, but *why* it is true.¹⁸ This is just Broad's point. He grants that even for emergent properties there are laws like (7). In case of emergent properties, however, these laws themselves cannot be explained any further, but are "unique and ultimate". So, we still have no answer to the question of what it means that systems having certain microstructures "are bound" to have certain properties or that certain properties of systems can be deduced from their microstructure.

A hint as to what such an answer could look like can, however, perhaps be found in the central passages of Broad's considerations on the concept of emergence. For it is a striking fact that in these passages Broad switches back and forth between talking of the behavior of an object and talking of its properties as if there were no difference. And this is, I think, no accident. It only reflects the fact that most (if not all) of the macroproperties

¹⁸ Cummins himself stresses this point when he writes:

"Instantiation laws are derived principles — or they *should* be in a full-dress theory — for they obviously call for explanation themselves. There is really only one available strategy for explaining an instantiation law: it must be derived from laws specifying the properties of the components. ... When we can do this, we can understand how P is *instantiated* in S ." (1983, p. 18)

If we take this into account, the schema of a "full-dress" property explanation according to Cummins looks like this:

- (8') (a) The properties of C_1, \dots, C_n are $\langle \text{whatever} \rangle$, respectively.
 (ii) S has microstructure $[C_1, \dots, C_n; R]$
-
- (iii) S has property F .

This, indeed, is almost identical with the schema (MR) developed below in this section except for two points. First, in general it will be impossible to deduce (iii) from (a) and (ii) if we do not also take into account the relevant laws which hold for objects with the properties specified in premiss (a), i. e., there is a premiss missing in the schema. Second, as will become clear in the remainder of this section, I do not think that it is necessary to show that sentence (iii) is deducible from sentence (ii) — by whatever means — if one wants to show that F is instantiated in S by $[C_1, \dots, C_n, O]$. The crucial point is that one can show that systems with microstructure $[C_1, \dots, C_n, O]$ have all the features characteristic of property F . That this does not amount to the same thing is especially shown by cases in which it is characteristic of F that systems having this property behave in a way describable by a law or lawlike sentence.

of objects are of a dispositional kind or can be analyzed in terms of the behavior of these objects. The color (or to be more precise: the reflectance) of a non-transparent object consists in its disposition to reflect electromagnetic waves of certain wavelengths and to absorb others. The liquidity or hardness of an object can be analyzed in terms of its tendency to change (or keep) its shape under certain conditions. Being magnetic can be analyzed in terms of tendencies to attract iron filings or to induce currents in certain kinds of conductors. To be sure, none of this implies that the predicates which express macroproperties can be defined in terms of behavior. It is only meant to imply that these predicates have very much the same status as the theoretical terms which have been widely discussed in the philosophy of science.

If T_C is a general theory of components such as C_1, \dots, C_n , one, therefore, can show that systems with microstructure $[C_1, \dots, C_n; R]$ "are bound" to have property F if one is able to show with the aid of T_C that systems with this microstructure behave in the way that is characteristic of property F .¹⁹ And this can be done *by deducing from T_C images of all laws which are characteristic of F* . (The set of these laws will be called L_F in the following.) In this context the image of a law of L_F is something very similar to what has been mentioned earlier: it is the law that results from such a law by replacing all occurrences of "has F " by "has microstructure $[C_1, \dots, C_n; R]$ ".

If this is so it is clear that there is a close resemblance between reductions in general and property explanations of this specific kind. Such explanations might as well, as I have already suggested, be called microreductions. But there is a remarkable difference which has not been addressed so far. In general we cannot assume that a certain macroproperty F can only be realized by one specific microstructure. Hence, in general we cannot assume that the predicate "has microstructure $[C_1, \dots, C_n; R]$ " is coextensive with the predicate "has F ". But if it is possible to deduce images of all elements of L_F from T_C , then this seems to imply that there must be such a coextensivity. At least, it seems plausible to assume that in such a case the expressions "has F " and "has microstructure $[C_1, \dots, C_n; R]$ " are nomologically coextensional, even if the corresponding nomological biconditional cannot be deduced from the original laws together with the image

¹⁹ Actually, this is only a special though important case. In general one has to show with the aid of T_C that systems with microstructure $[C_1, \dots, C_n; R]$ have whatever is characteristic of property F . If it is a certain kind of behavior, one has to show that systems with microstructure $[C_1, \dots, C_n; R]$ exhibit exactly this behavior. If it is a set of features F_1, \dots, F_n , one has to show that systems with microstructure $[C_1, \dots, C_n; R]$ have all these features. If it is something else X , one has to show that systems with microstructure $[C_1, \dots, C_n; R]$ have X .

laws. How can we avoid this unwanted result? I already mentioned the fact that in general macroproperties can be realized by a whole range of different microstructures. If property F is realized in system S_1 by the microstructure $[C_1, \dots, C_n; R]$, it may be realized in system S_2 by the microstructure $[C_1', \dots, C_m'; R']$, and in system S_3 by still another microstructure. What all these systems with all these different microstructures have in common is that they all behave in the way that is characteristic of property F . Microreduction thus in most cases is *system-relative*. In general we, therefore, cannot expect that it will be possible to deduce images of all elements of L_F from T_C . But that is not necessary either. For if we only want to show that system S_1 "is bound" to have property F if it has microstructure $[C_1, \dots, C_n; R]$, that is, if we want to show that in system S_1 property F is realized by this microstructure, it suffices to deduce from T_C for each law of the form "For all x : ... x has F —" of L_F the corresponding lawlike sentence "... S_1 has microstructure $[C_1, \dots, C_n; R]$ —".

If we call these lawlike sentences system-relative images of the elements of L_F we, thus, are finally able to define "microreduction" or "explanation" or "realization"²⁰ of macroproperty F by microstructure $[C_1, \dots, C_n; R]$ in system S as follows:

(MR) In system S the property F is microreducible to the microstructure $[C_1, \dots, C_n; R]$ iff there is a general theory T_C ²¹ from which a

²⁰ Some certainly would like to make a distinction between microreduction and microrealization, since the former, they will claim, requires biconditionals. But in my view this distinction is spurious. First of all, traditionally reducibility has not been understood as implying the existence of biconditionals. Take, e.g., the question of whether vital phenomena (breathing, digestion, procreation, goal-oriented behavior) can be reduced to the physical makeup of living bodies. Concerning this question nobody was ever interested in whether there are physical conditions P such that a living body x has the vital property F if and only if x is in P . Instead of this, people wanted to know whether we can see *how* certain physical conditions work together to bring about the effects which are characteristic of the vital phenomena in question. Second, even if one accepts the claim that microreductions presuppose the existence of biconditionals, microreductions will be no more than rather uninteresting special cases of microrealizations, cases in which as a matter of contingent fact there is only one microstructure that realizes this macroproperty. For in both cases what is really essential is whether we can deduce system-relative images of the relevant laws from a component theory T_C . (Cf. the definition (MR) below.)

²¹ Brian McLaughlin has argued in personal discussion that according to this definition the solubility of sodium chloride is not micro-reducible if one understands T_C literally as a theory of the components of sodium chloride only. For the fact that sodium chloride dissolves in H_2O depends, in part, also on laws governing hydrogen and oxygen which are not components of sodium chloride itself. I think, that in this case one could counter this objection by arguing that the relevant parts here are not

system-relative image can be deduced for each of the laws that are characteristic of F .²²

The term "general" in this definition is meant to account for Broad's considerations mentioned above. That is, a general theory T_C is meant to be a theory which relies only on the basic properties of the components C_1, \dots, C_n ²³ and which does not comprise "unique and ultimate" principles of the kind "Every object with microstructure $[C_1, \dots, C_n; R]$ behaves in the way xyz ". In my view, this so defined concept of microreduction gives a clear and reasonable explication of what Broad wanted to express by saying that a property can be deduced from the properties of its components and their manner of organization. But if this is so, the concept of microreduction is the exact complement of the concept of emergence. A macroproperty F is emergent if and only if it is not microreducible. So, the concept of emergence can perhaps best be defined like this:

- (E*) Let S be a system having the microstructure $[C_1, \dots, C_n; R]$, then F is an emergent property of S iff (a) there is a law to the effect that all systems with this microstructure have F , but (b) F is not microreducible to $[C_1, \dots, C_n; R]$.

4. Summary

If we compare the theories of supervenience and emergence as mind-body theories it is obvious that these two theories are of a similar kind. Both hold that mental properties are in a certain sense dependent on physical properties without being reducible to them. But the focal point of the two theories is somewhat different. Within the frame of the supervenience debate the emphasis has always been on the notion of dependence. Its aim was to find out whether there is a dependency relation weak enough to be compatible with the non-reducibility of the mental to the physical or,

at the level of atoms, but at the level of subatomic particles such as protons, neutrons, and electrons. And these certainly are parts of sodium chloride. But in general, I think, McLaughlin's point is well taken, and I, therefore, propose to understand T_C to be a theory not only of the components of the system under consideration, but of all entities which belong to the same level as these components.

²² Obviously there is a close resemblance between the notion of microreduction understood in this way and the notion of *local* reduction which Kim developed e. g. in (1982) and (1989). But there are also differences. For according to Kim local reduction rests on local bridge laws whereas bridge laws play no role in this account.

²³ And other entities which are at the same level as these components. Cf. note 19. above.

what would have been regarded as even better, with the non-existence of psychophysical laws or any other property-to-property connections between these two realms. But the result of its efforts were rather meager. Weak and global supervenience have been shown to be too weak to deserve the name of real dependency relations. So, only strong supervenience is left. And strong supervenience at least implies the existence of laws from the physical to the mental.

The notion of reduction, on the other hand, was never of great interest for those who took part in the supervenience debate. Most authors took it to be understood that a mental property F is reducible to a physical property G in exactly those cases in which it can be shown to be nomologically coextensive with G , perhaps with the additional constraint that G determines a natural kind in the realm of physics. At least for C. D. Broad, however, the notion of reduction was of utmost importance. He did not say very much about dependency, since for him it was clear from the beginning that the dependence of the mental upon the physical can only consist in the existence of laws connecting physical with mental properties. For otherwise the (emergent) mental properties could not be explained by appeal to the corresponding physical ones. So, it seems that Broad and the proponents of strong supervenience both agree on the following claim which is an expression of the notion of dependence they share.

- (D) If an object a has the mental property F there must be a (microstructural) physical property G such that
- (i) a has G , and
 - (ii) there is a nomological connection between G and F , that is, the sentence "For all x : if x has G , x has F " is a true nomological generalization.

It is the concept of reduction where roads part, since the theory of emergence claims that the pure existence of bridge laws is neither a necessary nor a sufficient condition for reducibility. It is not necessary since a mental property F can be microreducible even if there is no appropriate bridge law that connects F with a corresponding physical property G .²⁴ And it is not sufficient, since even if there is such a bridge law, F can fail to be microreducible.²⁵

²⁴ For in order to show that F is microreducible it suffices to deduce the relevant system-relative image-laws from an appropriate theory T_G . And this can be done without using any bridge laws at all.

²⁵ This, e. g., is the case if the bridge law itself is "unique" and "ultimate" in Broad's sense.

In my view, this analysis of reducibility by microreducibility is much to be preferred. If a property F is reducible to a property G , the possession of G must in a strong sense *explain* the possession of F . Hence, what we want to know is not *whether* there exists a lawful connection between F and G , but *why* it exists (if it does). Laws or even bridge laws do not yield such an explanation, rather they are statements of what needs to be explained. Microreductions on the other hand tell us what we want to know because they make it clear that objects with such-and-such a microstructure must behave in the way that is characteristic of a certain macroproperty. Microreductions, therefore, show why objects with this microstructure *must* have this macroproperty. Compared with the theory of supervenience the theory of emergence, thus, is the better explication of what could be meant by "physicalism without reductionism". But this of course does not show that the theory of emergence as a mind-body theory is true, rather it shows what the yard stick for a more ambitious physicalism would look like.

References

- Broad, C. D. (1925) *The Mind and Its Place In Nature*. London: Routledge and Kegan Paul.
- Burge, T. (1979) "Individualism and the Mental". *Midwest Studies in Philosophy* 4, pp. 73–121.
- (1986) "Individualism and Psychology". *Philosophical Review* 95, pp. 3–45.
- Churchland, P. M. (1985) "Reduction, Qualia, and the Direct Introspection of Brain States". *Journal of Philosophy* 82, pp. 8–28.
- Churchland, P. S. (1986) *Neurophilosophy: Toward a Unified Science of Mind|Brain*. Cambridge, Mass.: MIT Press.
- Cummins, R. (1983) *The Nature of Psychological Explanation*. Cambridge, Mass.: MIT Press.
- Davidson, D. (1970) "Mental Events", in: Foster, L., and Swanson, J. W. (eds.) *Experience and Theory*. Amherst: University of Massachusetts Press, pp. 79–101. Reprinted in: Davidson, D., *Essays on Actions and Events*. Oxford 1980, pp. 207–225.
- Fodor, J. A. (1974) "Special Sciences, or The Disunity of Science as a Working Hypothesis". *Synthese* 28, pp. 97–115. Reprinted in: Fodor, J. A., *Representations*. Cambridge, Mass.: MIT Press, 1981, pp. 127–145.
- Glöckner, W. (ed.) (1973) *Fischer Kolleg Chemie*. Frankfurt/M.: S. Fischer Verlag.
- Grimes, Th. R. (1988) "The Myth of Supervenience". *Pacific Philosophical Quarterly* 69, pp. 152–160.
- Hempel, C. G., and Oppenheim, P. (1948) "Studies in the Logic of Explanation". *Philosophy of Science* 15, pp. 135–175. Reprinted in: Hempel, C. G., *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York 1965, pp. 245–290.
- Hooker, C. A. (1981) "Towards a General Theory of Reduction". *Dialogue* 20, pp. 38–60, 201–236, 496–529.

- Horgan, T. (1982) "Supervenience and Microphysics". *Pacific Philosophical Quarterly* 63, pp. 29–43.
- Kim, J. (1982) "Psychological Supervenience as a Mind-Body Theory". *Cognition and Brain Theory* 5, pp. 129–147.
- (1984 a) "Supervenience and Supervenient Causation". *The Southern Journal of Philosophy* 22. Supplement. *Spindel Conference 1983: The Concept of Supervenience in Contemporary Philosophy*, pp. 45–56.
 - (1984 b) "Concepts of Supervenience". *Philosophy and Phenomenological Research* 45, pp. 153–176.
 - (1987) "'Strong' and 'Global' Supervenience Revisited". *Philosophy and Phenomenological Research* 48, pp. 315–326.
 - (1989) "The Myth of Nonreductive Materialism". *Proceedings and Addresses of the American Philosophical Association* 63, pp. 31–47.
 - (1990) "Supervenience as a Philosophical Concept". *Metaphilosophy* 21, pp. 1–27.
- Nagel, E. (1961) *The Structure of Science*. London: Harcourt, Brace and World.
- Putnam, H. (1975) "The Meaning of 'Meaning'", in: Gunderson, K. (ed.), *Language, Mind, and Knowledge, Minnesota Studies in the Philosophy of Science, Vol. 7*. Minneapolis: University of Minnesota Press. Reprinted in: Putnam, H., *Mind, Language, and Reality, Philosophical Papers, Vol. 2*. Cambridge: Cambridge University Press, pp. 215–271.
- Teller, P. (1984) "Comments on Kim's paper". *The Southern Journal of Philosophy* 22. Supplement. *Spindel Conference 1983: The Concept of Supervenience in Contemporary Philosophy*, pp. 57–61.