

## **Emergent Properties**

Timothy O'Connor

*American Philosophical Quarterly*, 31 (1994), 91-104

All organised bodies are composed of parts, similar to those composing inorganic nature, and which have even themselves existed in an inorganic state; but the phenomena of life, which result from the juxtaposition of those parts in a certain manner, bear no analogy to any of the effects which would be produced by the action of the component substances considered as mere physical agents. To whatever degree we might imagine our knowledge of the properties of the several ingredients of a living body to be extended and perfected, it is certain that no mere summing up of the separate actions of those elements will ever amount to the action of the living body itself.

– J.S. Mill, A System of Logic, Bk.III, Ch.6, §1

In attempting to develop an ontology adequate to account for some of the more puzzling features of the natural world, several philosophers and scientists in the past century have tried to articulate a via media between the extremes of radical dualism and reductionism. This middle road consists in the claim that the phenomenon in question is at once grounded in and yet emergent from the underlying material structure with which it is associated. At various times, this claim has been made with respect to the so-called secondary qualities, biological life, and, most commonly of all, consciousness. In spite of its obvious attractions as a way to reconcile apparently conflicting data (both philosophical and empirical), the emergentist strategy is regarded with grave suspicion by many contemporary philosophers. Clearly one reason it has been viewed in this way is that various formulations of the very notion of 'emergence' have been imprecise and not obviously reconcilable with one another. And even more worrisome for many is the seeming violation of the maxim that you can't get something from nothing.

My central purpose in what follows is to bring greater clarity to this whole discussion, a task of some importance in view of the modest renewal of interest in emergentist hypotheses about the mental in recent days. I begin (in section I-II) by canvassing some of

the major formulations of the concept of property emergence and pointing out their weaknesses, before providing an alternative account (in section III). My criticisms of these earlier approaches do not take the form of charging that they fail to accord with a pre-analytic notion, since whatever content has come to be associated with this term through previous philosophical usage is sufficiently vague as to allow for diverging developments. Rather, these other accounts are inadequate because they fail to capture the sort of macro-micro relationship that would best fit the desiderata suggested by reflection on the puzzling phenomena that motivate the enterprise in the first place. Thus, my remarks amount to a suggestion concerning how to elucidate the notion of 'property emergence' in a way that best fits the purposes of philosophical and scientific theory.

After setting out my preferred account of property emergence and responding to some possible objections to it, I try (in section IV) to clarify the notion of emergence operative in the theory of human consciousness that the neurophysiologist R.W. Sperry has articulated over the past couple of decades.

## I

An influential early account of emergence is found in Samuel Alexander's Space, Time, and Deity (1920)<sup>1</sup>. It will be convenient to begin the examination of that account with the following passage, which readily lends itself to the way that Alexander is typically understood:

Physical and chemical processes of a certain complexity have the quality of life. The new quality life emerges with this constellation of such processes, and therefore life is at once a physico-chemical complex and is not merely physical and chemical, for these terms do not sufficiently characterize the new complex which in the course and order of time has been generated out of them. Such is the account to be given of the meaning of quality as such. The higher quality emerges from the lower level of existence and has its

roots therein, but it emerges therefrom, and it does not belong to that level, but constitutes its possessor a new order of existent with its special laws of behaviour. The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I should prefer to say in less harsh terms, to be accepted with the "natural piety" of the investigator. It admits no explanation. (pp.46-7)

The usual way of reading remarks such as these seems to go something like this: once a material object or system attains a sufficient level of structural complexity, it thereby, as a matter of nomological but not metaphysical necessity, acquires an emergent quality, Q. Q importantly differs from typical macro-properties like, e.g, the shape of a table, which seems to be appropriately describable as a mere "resultant" of properties of the object's micro-structure.<sup>2</sup> That difference is reflected in the fact that it introduces a change in the behavior of the object which possesses it, such that processes internal to that object cannot be explained in terms of the laws which govern its micro-particles in the absence of such organizational complexity.

We will see aspects of this sort of view in some of the accounts to be discussed later, but it clearly misrepresents Alexander's own view. Both in the central discussion of emergent qualities (Bk.III, Ch.II) and in his later discussion of the nature of human freedom of action (Bk.III, Ch.X), Alexander insists that every event can be given a deterministic explanation at the level of the world's fundamental spatiotemporal features, an explanation devoid of reference to emergent qualities. A LaPlacian calculator who had a full knowledge of the laws of fundamental physics and the total distribution of matter at some point during the nebular period, e.g., could predict, under a microphysical description, the event which is my presently sitting in this chair and pondering the nature of emergent properties.

On the other hand, Alexander also claims that such a predictor would fail to give a fully sufficient characterization of this event. Furthermore, where her knowledge is limited in the

way I've described, she couldn't adequately characterize that future event, owing to an inability to anticipate and take due account of the presence of the relevant emergent qualities. And this, despite the fact that the event which is characterized in terms of fundamental spatiotemporal features is identical to the event described at the macro-level with reference to those emergent qualities.

Now one might try to reconcile all this with the standard interpretation of Alexander's views as I've represented it by concluding that Alexander is committed to levels of existence within one reality that are fully harmonized and yet (in a very strong sense) irreducible in that there is no causal influence across levels; rather, each level has, as Alexander himself says, its own "methods of behavior", its own unique, fully adequate set of laws governing processes at that level. (To forestall what might seem to be an obvious objection, let me add that macro-qualities of objects that are not emergent, but only "resultant", such as mass, do not belong to the 'higher levels' of that object's existence as Alexander conceives them.)

Such a reading, however, would only serve to make Alexander's view irremediably mysterious. If fundamental physical laws are sufficient to fix the micro-physical features of any event, then how could an object's lower level micro-structure fail to ultimately determine in some manner its behavior at the higher level? This total bifurcation of levels is far more difficult to make sense of than the dichotomy posited by ordinary substance dualism, since it is held to occur within a single spatiotemporal manifold, and, indeed, within the very structure of individual physical objects.

And, in any case, the standard interpretation cannot be maintained once one recognizes that Alexander identifies emergent qualities with (highly general) 'configurational' patterns. This is suggested, e.g., by a remark which immediately follows the striking passage with which we began:

To adopt the ancient distinction of form and matter, the kind of existent from which the new quality emerges is the 'matter' which assumes a certain complexity of

configuration and to this pattern or universal corresponds the new emergent quality.

(p.47)

(He goes on to say that this correspondence is one of identity.) I believe that Alexander's idea of 'configurational pattern' is best understood in terms of the contemporary notion of a structural property. I suggest that we define this notion in the following manner:

A property, S, is structural if and only if proper parts of particulars having S have some property or properties not identical with S, and this state of affairs is, in part at least, constitutive of the state of affairs of the particular's having S.<sup>3</sup>

To get a rough and ready handle on this notion, let us momentarily put to one side the status of an emergent property and instead broaden our focus to consider the implications of a certain way of looking at the ontology of properties more generally. Suppose you hold (with Alexander) that there are universals which are shared by various particulars, and that a given particular's having a universal (my shirt's being blue) is not a matter of its being related to a transcendent abstract object (the Form of Blueness), but rather consists in that universal's being wholly present right there in the object itself. If so, one might reasonably be inclined to take it that (at least as a general matter) the having of properties by complex, macro-physical objects is not anything 'over and above' the having of all the various micro-physical properties and relations by the parts of that same object. (One such reason comes from the intimate connection ordinarily posited between properties and causal powers. Ordinary macro-physical properties such as having a particular determinate chair-like shape do not add any causal potentialities to objects that have them exceeding the 'summation' of potentialities afforded by the micro-properties and relations acting in concert.) And if not anything 'over and above' the micro-properties, then such (instantiations of)<sup>4</sup> macro-properties might naturally be thought to be constituted by them.

With this rudimentary grasp on the notion of a 'structural property', we may now return to Alexander's characterization of an emergent. If an emergent quality is in this way a property-structure constituted by 'lower-level', non-emergent properties, and, therefore, presumably no different from a mere 'resultant' macro-property (such as an object's shape) in terms of its fundamental nature, it cannot (as the standard reading would have it) in any way introduce a pattern of behavior at the fundamental level differing in kind from the sort that would occur in its absence. If its nature is fully fixed by the properties out of which it is constituted, then so must be its role in determining the causal processes involving the object which bears it.

How, then, are we to make sense of Alexander's view? How may we reconcile, on the one hand, the claim that emergent qualities cannot be predicted on the basis of micro-structure and involve "special laws of behaviour" which "admit no explanation" (p.46-7) with the identification of emergents with certain configurations of features of the micro-level, on the other?<sup>5</sup> I suggest that we think of the matter in the following way. Emergents, on Alexander's view, clearly enough are a special sort of complex structural properties.<sup>6</sup> Apparently, one important feature is resistance to dissolution, as exhibited at the level of chemical molecules and again at the level of organic life, with its processes of self-regulation. But the true criterion, I think, is the availability of laws fully governing the behavior of entities at that level. Consider the following remarks:

If. . . the processes of a particular level are represented as a processes, a constellation of such processes is of such a kind as to be a new process ab with its quality B. That is, the thing which is based on that constellation of a processes has an emergent quality B, whose behaviour consists in ab processes; and though ab processes are also a processes they are not merely such, and are on a different level from the processes which are sufficiently distinguished from other forms of existence as being merely a processes. (p.46, emphasis added)

As I understand it, the claim here is that emergent qualities are involved wherever we have processes which admit of complete explanation at two or more different levels of complexity (as represented by his a and b processes). But why, then, does he also claim that emergent qualities - which, again, are nothing but 'configurations' of micro-properties - cannot be predicted on the basis of micro-level laws alone, and that the existence of laws involving them “admit no explanation”?

A clue is provided in another passage:

To call [a structure] organism is but to mark the fact that its behaviour, its response to stimulation, is, owing to the constellation, of a character different from those which physics and chemistry are ordinarily concerned with, and in this sense something new with an appropriate quality, that of life. At the same time, this new method of behaviour is also physico-chemical and may be exhibited without remainder in physico-chemical terms, provided only the nature of the constellation is known - provided, that is, we remember. . . that there is already a constitution in the organism. . . . Until that constellation is known, what is specially vital may elude the piecemeal application of the methods of physics and chemistry. (p.62)

I suggest that these remarks, when placed alongside the previous passage, are best interpreted as follows: one who has a complete grasp of the fundamental laws of physics will not thereby have a basis on which to predict the special scenarios under which material processes will admit of a wholly autonomous form of explanation making reference only to higher-level structural properties. The completely general laws take no account of special-case (emergent) scenarios, with their attendant special-case laws. Alexander seems to recognize at the same time that once our LaPlacian predictor's attention is drawn to the special natures of chemical, organic, and conscious entities, she could in principle go back

and derive the laws of chemistry, biology, and psychology, respectively, from those of physics.<sup>7</sup> Finally, it may be that Alexander is claiming that the existence of emergent levels allowing for unique forms of explanation itself “admits no explanation” either because we cannot say why the fundamental laws happen to be such as to yield such special case forms, or because we cannot explain why the universe’s initial conditions were such as to allow for the realization of the potential for these special forms of laws, or perhaps both.

With the essential features of Alexander's account of emergence before us, we can see that it is a much weaker view than some of his language might suggest. On this view, there is no intrinsic difference between emergents and garden-variety supervenient structural properties; it's just that properties of the emergent sort fall under a plurality of distinct groupings of causal laws. Since these properties introduce no novel causal influence over the behavior of objects bearing them, these 'distinct' groupings of laws (physical, chemical, biological, and psychological) are not ultimately independent of each other, even in part; laws pertaining to higher levels of organizational complexity are simply special cases of the more fundamental levels. Alexander has thus failed to capture a sufficiently rich understanding of the notion of 'emergence', one, e.g., that would provide a conceptual framework permitting a middle road between substance dualism and various traditional forms of materialism. Therefore, despite the fact that he has played an important role in getting emergentist views into philosophical discussion, we will need to look in other directions to find a clear and useful account of the fundamental underlying concept.

A second early exposition of the notion of emergence that I want (more briefly) to discuss is Arthur Lovejoy's essay, "The Meaning of 'Emergence' and its Modes" (1927).<sup>8</sup> He actually considers a variety of possible forms of emergence, such as the emergence of natural laws or new sorts of fundamental particulars. Here I will simply confine my discussion to our target notion of property emergence. Lovejoy suggests that we can get at the notion of emergence by imaginatively juxtaposing two phases of the universe's history. (Call the earlier of these "phase A", and the later one "phase N".) On his view, emergence

has occurred if there are event- or process-types in phase N that are irreducibly different in kind from any present in A (p.591).

Events, as I will be thinking of them, are exemplifications of properties by one or more objects at a particular time. (The adoption of any other current account of events which construes them as particulars would not affect the remarks which follow in any essential way.) An event which is irreducibly different in kind from any of a certain set of events, therefore, would seem to require the instantiation of a fundamentally new type of property, i.e., one differing in kind from any of the structural properties or (presumably nonstructural) fundamental properties exemplified by members of that set. (An event which differed from some others only in terms of its level of complexity would not be irreducibly different, since it would be a mere amalgam of property instantiations of a sort exhibited by at least some of those other events.) The defining features of the sort of property required, then, are: that it is (a) potentially had only by objects of some complexity, (b) not had by any of the object's parts,<sup>9</sup> and (c) distinct from any structural property of the object. For convenience, we may assume that all such properties are simple in nature, but it should be noted that it is also possible that there be complex instances as well, where the complexity would consist in its being a conjunction of two or more such (same-level) simple properties.<sup>10</sup> (The reader will recall that a structural property, by contrast, is constituted by properties of the object's parts.)

The features I have just specified clearly are central components of what many philosophers have in mind in employing the notion of emergence. The thesis that consciousness is an emergent phenomenon, e.g., commonly is intended to express a claim entailing that it introduces a qualitatively new (macro-level) feature into the world. However, to take this as a sufficient condition for property emergence is to countenance the possibility of causally idle, i.e., epiphenomenal, emergents, which is undesirable in view of the purposes to which the notion of an emergent property is typically put. Lovejoy's additional mention here of fundamentally new sorts of processes as well as events is pretty clearly intended to

point to emergent properties that go beyond playing a purely epiphenomenal role, since a causally-connected series of events of a qualitatively new character within an object or group of objects will perforce involve the causal influence of one or more emergent properties.

## II

Our investigation thus far, then, suggests that an emergent property is best thought of as a (typically) simple, non-structural natural property which is exemplified by objects or systems that attain the appropriate level and kind of organizational complexity and which exerts a causal influence on the behavior of its possessor. The task that remains is to clarify this idea with respect to (i) the nature of the dependency of the emergent property upon the lower level properties of the object and (ii) the general nature of an emergent's causal influence.

One recent attempt to tackle the first part of this problem is James van Cleve's "Mind-Dust or Magic? Panpsychism versus Emergence" (1990).<sup>11</sup> Van Cleve suggests that we think of emergence as a species of what Jaegwon Kim terms "strong supervenience", which is captured in the following<sup>12</sup>:

A-properties supervene on B-properties = Df. Necessarily, for any object x and A-property a, if x has a, then there is a B-property b such that (i) x has b, and (ii) necessarily, if anything has b, it also has a.

("A-properties" and "B-properties" here refer to families of properties. In the case of emergents, the A-properties will be the set of emergent properties and the B-properties may be understood as the set of (non-emergent) physical properties generally.) As van Cleve notes, this definition is intended to capture (a form of) supervenience of a property of an object on some other property of that very same object. Emergent properties, however, are ordinarily characterized in relation to properties of the object's parts. (In the case of the

thesis of emergent consciousness, e.g., the subvening properties are most plausibly taken to be neural structures.) It is easy enough, though, to modify the definition to apply to such cases:

A-properties of objects supervene on B-properties of their parts = Df. Necessarily, for any object x and A-property a, if x has a, then there are B-properties b,c,d, . . . (including relational properties)<sup>13</sup> such that (i) some proper parts of x have (variously) b, c, d, . . . and (ii) necessarily, for any things collectively having all of b, c, d, . . . there is an object of which they are parts that has a.

Now both of these definitions suggest that the notion of supervenience encompasses two basic ideas, which van Cleve helpfully characterizes in the following manner<sup>14</sup>:

The first component, corresponding to clause (i), is dependence on the B-properties: nothing can have A-properties unless it [or some part of it] also has B-properties. The second component, corresponding to clause (ii), is determination by the B-properties: nothing can be just like a given thing as regards its [or its parts'] B-properties without also being just like it as regards its A-properties. (p.221)

These two features of dependence and determination, then, enable us to clarify a way in which certain properties of an object are modally connected to the properties of its parts. This applies to unproblematic examples such as an object's mass or shape as well as (according to the present suggestion) to our target notion of emergence. How should we distinguish emergent properties from their less elusive congeners?

Van Cleve takes his cue from some remarks of C.D. Broad (1925), who states that an emergent property is to be inferred wherever

the characteristic behavior of a whole could not, even in theory, be deduced from the most complete knowledge of the behavior of its components, taken separately or in other combinations, and of their proportions and arrangements in this whole.<sup>15</sup>

On Broad's view, then, an emergent property of a whole may be distinguished from garden-variety supervenients like mass in virtue of the fact that the former (but not the latter) sort of property could not be deduced from a complete knowledge of the properties of its components. Van Cleve suggests that what Broad has in mind here may be captured by restricting the form of necessity possessed by the inner modal operator in our definition above. Since ordinary supervenients are deducible from the properties of the object's parts, having the right sort of subvening properties is logically sufficient for having the supervenient property. Emergent properties, however, follow from the subvening properties only with nomological necessity (i.e., they obtain under the associated subvening conditions only in worlds with the same (contingent) causal structure as our own). Hence,

If P is a property of w, then P is emergent iff P supervenes with nomological necessity, but not with logical necessity, on the properties of the parts of w. (p.222)

There are two basic problems with this way of characterizing emergence as a species of supervenience. The most serious of these is that it counts a wide range of properties as emergents that quite clearly do not belong to that category. For his account of emergence critically depends on the view that causal or nomological necessity is weaker than metaphysical or broadly logical necessity. But if that view is right, it would seem plausible to suppose that, e.g., the potentiality of a knife to slice a loaf of bread is a macro-property which supervenes on the properties of its molecular constituents with nomological necessity only. (If facts about the causal structure of our world are in general logically contingent, it is hard to see why this sort of supervening potentiality would be connected to its underlying

physical basis in a stronger fashion.) Examples of this sort could be multiplied indefinitely. (Just pick out a causal potentiality of an object that is closely associated with its macro-structure.) Clearly, though, the distinction the weak necessity view allows us to draw between this sort of property and simple structural properties which are logically entailed by the underlying micro-properties is not the sort emergence theorists typically have in mind.

A second problem with van Cleve's explication of property emergence is its very reliance on a controversial understanding of causal necessity. It is an old and respectable view (if currently out of favor) that it is part of the nature of a natural property of an object that it have the potential for contributing to certain characteristic effects. This implies that the relationship between an object's properties and its causal powers is a logically necessary one.<sup>16</sup>

Now, whether this understanding of the relationship between an object's properties and its causal powers is correct (as I believe) or not, it is implausible to suppose that a proper analysis of emergence will imply its falsity. Yet this is a consequence of van Cleve's account, since he needs to draw a distinction between two classes of supervenients, such that in one the members follow logically from their associated subvening properties and in the other the connection is causal, but logically contingent. It clearly is preferable that we be able to give an account of emergence that does not rest on such a major assumption concerning the nature of causal necessity.

### III

To provide such an account, we need only to fit together the various elements that have already been suggested. These features are supervenience, non-structurality, and novel causal influence. It is vital to a satisfactory account of property emergence, I think, that it incorporate all of these conditions.

By contrast, recent authors who have attempted to explicate the notion of emergence have done so exclusively in terms of one or another of these. Van Cleve, as we have seen,

suggests that emergence is definable in terms of the way in which a property supervenes on the properties of an object's parts. Armstrong (1978) focuses solely on non-structurality. Yet a third author, Robert Klee (1984)<sup>17</sup>, draws upon the idea of causal influence. I have already criticized van Cleve's account. Against both Armstrong and Klee, I would urge that supervenience is needed if emergence is to be capable of being incorporated within a scientific framework. Without the two components of determination and dependency, there would be no potential for uncovering precise causal conditions under which emergence occurs.

Recall, then, the part-whole variation on Kim's account of (a form of) supervenience that I suggested earlier:

A-properties of objects supervene on B-properties of their parts = Df. Necessarily, for any object *x* and A-property *a*, if *x* has *a*, then there are B-properties *b, c, d, . . .* (including relational properties) such that (i) some proper parts of *x* have (variously) *b, c, d, . . .* and (ii) necessarily, for any things collectively having all of *b, c, d, . . .* there is an object of which they are parts that has *a*.

This strong form of supervenience is well-suited to an account of property emergence.<sup>18</sup> Contrary to van Cleve, however, there is no need to suggest that the form of necessity captured by the inner modal operator is weaker than that appropriate to non-emergent supervenients. It does seem right to say that the relevant form of necessity is causal, given that the occurrence of an emergent property is presumably a function of the causal potentialities of underlying base properties. But since emergence cannot be satisfactorily differentiated from other forms of supervenience in terms of the logical strength of the connection between such a property and its underlying base, one is free to understand the operator in the account which follows in terms of one's favored view of causal necessity, whether it be as a species of metaphysical necessity or as a sui generis and logically contingent type.

Our second feature of emergence, which I am calling "non-structurality" for lack of a better term (again, this is distinct from simplicity or non-complexity), itself involved three components: the property's being (a) potentially had only by objects of some complexity, (b) not had by any of the object's parts, (c) distinct from any structural property of the object.

Finally, there is the idea of "novel causal influence". This term is intended to capture a very strong sense in which an emergent's causal influence is irreducible to that of the micro-properties on which it supervenes: it bears its influence in a direct, "downward" fashion, in contrast to the operation of a simple structural macro-property, whose causal influence occurs via the activity of the micro-properties which constitute it. Of course, if we take emergents to be a species of supervenient properties, as I have suggested, then the continuing instantiation of the emergent property is completely dependent upon some set of properties or disjunctive ranges of properties in the object's microstructure. Nonetheless, it exerts a causal influence on the micro-level pattern of events that is not reducible to the immediate causal potentialities of the subvening properties.<sup>19</sup>

Employing the three component notions just canvassed, then, I offer the following definition of property emergence:

Property P is an emergent property of a (mereologically-complex) object Q iff:

- (1) P supervenes on properties of the parts of Q; and
- (2) P is not had by any of the object's parts; and
- (3) P is distinct from any structural property of Q, and
- (4) P has direct ("downward") determinative influence on the pattern of behavior involving Q's parts.

Having laid out the basic contours of the account, we must pause to consider a powerful objection to it.<sup>20</sup> If an emergent property is a necessary consequence of certain base-level properties (as is implied by the supervenience condition), then its instantiation is one of the potentialities of that set of properties. But then are not the further potentialities of this emergent property also a subset of the total set of potentialities of the base properties, in virtue of the necessary connection between the base properties and it? These further potentialities are simply potentialities of the base properties at one remove. And now one is led to wonder why we might ever think to postulate an emergent property at all, since it provides no explanatory gain over an account which excises the mediating link by taking the "further" potentialities as directly tied to the base properties.

This objection implies, in effect, that the features of supervenience and novel causal influence are incompatible. That it is mistaken, however, can be seen once we have more precisely characterized the sense in which potentialities of an emergent property "go beyond" those of the base properties. As a simplifying assumption, suppose there were only one naturally-emerging property, P, and that P is instantiated only in systems of a rather high level of complexity ("n"). Suppose further that physicists had come to an understanding of a set of (fundamental level) laws, L, that accurately described the processes of matter for all systems whose levels of complexity were lower than n, but failed fully to govern these complex systems of level n. In such a scenario, there would be good reason to surmise that here we had an emergent property (or properties) at work. But since the property would have (we may suppose) various effects in a lawlike way (as ordinary, non-emergent properties do), there seems to be no bar in principle to our physicists' revising their formulation of the fundamental laws to take account of this quirky phenomenon. And, hence, it is true (as is hinted in the objection of the previous paragraph) that, owing to the necessary connection between the base properties and P, even the quirky phenomenon could be described in terms of functions from the base-level properties alone. But this does not

motivate the repudiation of the presence of emergent properties. For the laws adequate to describe the quirky phenomenon will themselves have a very odd complexity, involving tacked-on disjuncts to cover the special cases. And this, surely, demands explanation in terms of the properties of the object exhibiting the strange behavior, an explanation that the postulation of an emergent property seems to provide.<sup>21</sup>

As Sydney Shoemaker has pointed out to me, however, there will always be an alternative explanatory possibility available to the theorist determined to "go it alone" without emergentist hypotheses. And that is to posit the presence of further (hitherto undetected) micro-properties. While I have no argument to show that one cannot (in consistency) make such a move, it is, in my judgment, an implausible one. Why does such a micro-property make its presence known only in highly complex systems of a certain sort? How is it that such a fundamental property can be so causally isolated from other micro-properties so as to be discernible only in circumstances that are otherwise noteworthy only for the complex macro-properties which are instantiated? The presence of an emergent property is by far the more natural assumption to make in the idealized circumstance depicted above, and the only motivation one could have for postulating a (rather elusive) micro-property is a very strong methodological principle to the effect that one is to avoid emergentist hypotheses at all costs, which by my lights is not a reasonable one.

As an alternative to the claim that there are further micro-properties, one might maintain that the micro-properties there are have causal features that don't come into play until the particles having the micro-properties become parts of systems having certain structures and degrees of complexity.<sup>22</sup> But, if I am understanding this idea properly, it seems to involve a micro-particle's having different effects in the same sort of local situation, depending on the broader context in which that local situation is itself imbedded. But here we avoid the sort of 'downward' causal influence that the emergentist hypothesis envisages only at the cost of a causally inexplicable 'responsiveness' of micro-level behavior to macro-level circumstances.

But, you may ask, isn't the emergentist, too, committed to micro-properties having causal potentialities that only come into play under the right macro-circumstances: the very potentiality to give rise to an emergent property under certain circumstances? So why not just say, in relation to your scenario, that there is only the one, non-emergentist sort of 'hidden' potentiality and be done with it?<sup>23</sup>

Here, we must make a distinction. It is true that the emergentist will suppose there to be a potentiality attached to a range of micro-properties that is realized only under special scenarios at the macro-level: the potentiality to give rise to an emergent in conjunction with various other micro-properties and relations. However, what the objector is committed to supposing is that a given micro-property could begin to have different characteristic effects in its local situation in 'response' to a macro-situation, a response that is ex hypothesi causally inexplicable. That, I say, is a mystery we do well to avoid.

Another, frequently-voiced objection to the notion of property emergence is that, even if one grants the internal coherence of the sort of account I have sketched, the presence of irreducible emergent properties in material substances is incompatible with contemporary scientific knowledge.<sup>24</sup> (A weaker variation on this claim would be that we know, at any rate, that there are no such properties associated with normally-functioning human brains.)

My reply to this bold, "tough-minded" claim is to label it as sheer bluff. If, as it seems to be the case, we cannot deduce a priori that no complex material substances have emergent properties, then the establishment of the "scientific knowledge" claim requires evidence drawn from the various studies of complex physical systems, particularly those falling under the purview of the biological sciences. And while it may be readily acknowledged that there are no widely-accepted working theories that are committed to the existence of such properties,<sup>25</sup> it seems clear enough that contemporary scientific knowledge is sufficiently partial as not to rule out emergentist hypotheses. (One need only consider, for example, the recent, proposals of macro-determinative influence on lower-level sub-structure by Polanyi<sup>26</sup> and Sperry<sup>27</sup> with respect to embryonic cells and consciousness, respectively.<sup>28,29</sup>

Whatever the merits of these conjectures may be, they support my outsider's contention that the body of firmly-established fact in the biological disciplines is capable of countenancing the possibility of property emergence.) This is perhaps especially so with regard to the complex workings of the human neurophysiological system. Given that so much remains to be understood about the detailed interactions of parallel and hierarchically-ordered subsystems of this system, how can it be confidently asserted that a completely general "bottom up" picture of this system is empirically established? We may wholeheartedly concur with Kim's assessment (1984, p.176) that micro-reduction has been an "enormously productive research strategy. . .in modern theoretical science," while insisting that current achievements are far from giving us conclusive grounds for asserting the general negative thesis that there are no macro-determinative emergent properties in nature.

Finally, I will simply point out that the account of emergence that I have presented in no way presupposes general causal determinism or its denial, and it thus implies the falsity of Karl Popper's contention that "the emergence of hierarchical levels or layers, and of an interaction between them, depends upon a fundamental indeterminism of the physical universe" (1977, p.35)<sup>30</sup>. Believing that unpredictability is the principle criterion of emergence, Popper notes that a sufficient degree of physical indeterminacy could make certain actual evolutionary developments involving the appearance of complex biological systems exceedingly improbable on remote prior conditions, and so unpredictable in principle. But, as we have seen, a stronger and, for certain purposes, more interesting understanding of emergent properties is available that allows that such emergents are strictly determined by the micro-properties on which they supervene. Furthermore, Popper has failed to see that emergents of the sort I have described can nonetheless retain an aspect of unpredictability: it would be impossible for an observer having a completely adequate understanding of the behavior of matter at levels of complexity below that which gives rise to emergents to fully predict its macro-properties and behavior when organized at that further level.

This completes my exposition and defense of the basic concept of emergence. However, I would like to call attention to just one of several difficult questions that could confront certain attempts to incorporate this notion into a developed theoretical account of specific natural phenomena: Is it possible for indeterministic processes to emerge at a certain level,  $n$ , in a system or structure from ‘determinism’ at levels below  $n$  (where what is meant here is that non-probabilistic laws are sufficient to fully describe the behavior of matter at levels below  $n$ )? This will be relevant to the attempt to give an emergentist account of human freedom of action, if one is inclined to hold (quite reasonably) that causal indeterminism is a necessary condition on freedom of action. For, although there is strong empirical evidence that nature is at bottom indeterministic, it is not so clear that indeterminism of the right sort can be generated via a direct function from quantum indeterminacy. For now I must leave this as a matter for future inquiry, one which I hope to address elsewhere.<sup>31</sup>

#### IV

I now will consider, finally, the “emergentist-interactionist” account of human consciousness that has been proposed by the distinguished neurophysiologist, R.W. Sperry, over the past two decades. Despite numerous publications by Sperry, there remains a great deal of controversy over just how to interpret his claims.<sup>32</sup> My hope is that the previous discussion of various concepts of emergence will help to clarify the precise content that Sperry attaches to the term “emergence”, and so serve to make clearer Sperry’s position on the nature of the mind as well.

Throughout his writings on the nature of human consciousness, Sperry repeatedly insists that, on his view, consciousness is an altogether new property relative to non-conscious matter, exerting a form of “emergent” or “downward causation” that is in no

sense reducible to the causal activity of the organism's component parts (see, e.g., 1991, p.222, p.236; 1986, p.267). Moreover, he suggests that in addition to the 'downward' influence of consciousness on underlying microstructure, it "determines the interactions of an entity as a whole at its own level" via "sequential same-level action" (1991, p.225; cf. 1980, p.200). Finally, his view also seems to incorporate the idea that consciousness is supervenient on neurophysiological properties:

. . . every time the elements of creation. . . are put together in the same way under the same conditions, [ ] the same new properties would emerge. . . (1980, p.200).

In short, the various passages to which I have referred suggest that Sperry is thinking of consciousness as an emergent property in pretty much the sense that I articulated earlier.

However, Sperry says a couple of other things about the relation of an emergent's causal activity relative to that of its underlying substrate that have led a number of commentators to suppose that the above remarks cannot be taken so straightforwardly, or to declare that Sperry has failed to enunciate a coherent thesis.

First, he repeatedly insists (e.g., 1980, p.201,202; 1986, p.268; 1991, p.225, 229 , 230-1, 240) that the "new laws" of emergentist interaction, as he conceives them, involve no violation or alteration of the laws governing elementary particles. Responding to criticisms of his account, Sperry asserts that

[t]he expectation that downward macrodetermination should thus effect reconfigurations. . . in the neuron-to-neuron activity of subjective mental states – or in the micro components of any macro phenomenon – indicates a serious misunderstanding of what emergent interaction is. From the start I have stressed consistently that the higher-level phenomena in exerting downward control do not disrupt or intervene in the causal relations of the lower-level component activity. Instead

they supervene in a way that leaves the micro interactions, per se, unaltered. (1991, p.230).

The second general claim that prompts some to give a watered-down interpretation of Sperry's view is his contention that emergent properties and causation are "ubiquitous, almost universal" in nature (1991, p.225; see also 1986, p.266), in no way confined to certain sophisticated biological organisms, but equally replete within the provinces of physics and chemistry (1991, p.229). Many of Sperry's critics (interestingly enough, these are primarily philosophers) have reasoned that since there is no warrant for supposing that emergents in the strong sense I outlined earlier are needed to deal with phenomena at the level of physics, in any case, the concept Sperry has in mind must in fact be ultimately reconcilable with a 'bottom-up' micro-determinism.

I think this is a fundamental misunderstanding of Sperry's position. That it is fairly widespread is to be explained by the highly novel character of Sperry's concept of emergence, which plays a critical role within an interesting, but difficult and perhaps imprecise, general philosophy of nature. This general thesis is made most explicit in the 1986 and 1991 articles. Immediately following the passage I quoted a couple of paragraphs back, in which Sperry stresses that the sort of downward causal control he envisages does not in any way disrupt the causal relations of the lower-level components, he adds:

These micro interactions and the interrelations of all the infrastructural components become embedded within, enveloped, and as a result are thereon moved and carried by the property dynamics of the larger overall system as a whole. . . that have their own irreducible higher-level forms of causal interaction.

A molecule within the rolling wheel, for example, though retaining its usual inter-molecular relations within the wheel, is at the same time, from the standpoint of an outside observer, being carried through particular patterns in space and time determined

by the over-all properties of the wheel as a whole. There need be no “reconfiguring” of molecules relative to each other within the wheel itself. However, relative to the rest of the world the result is a major “reconfiguring” of the space-time trajectories of all components in the wheel’s infrastructure. (1991, p.230)

This passage may helpfully be compared with the following (from the same article):

. . . except for ultra simple systems, we can never know “the relations existing between” the components. These include the enormously complex spacing and timing, pattern and form factors that involve the multi-variate, hierarchic, often dynamic infrastructures, and which, in themselves have causal control influence. Neither present-day science nor mathematics can handle these spacetime factors. No existing laws or principles are adequate. (p.235)

The latter passage strongly suggests that emergent properties of complex systems play a major role in determining the overall relational structure in which the component molecules are embedded. So in what sense do they not effect a “reconfiguration” of these components, as the earlier passage claims? I think Sperry’s idea here is that the presence of an emergent property and the resulting new relational structure are instantiated simultaneously, and are therefore not related as an efficient cause to an effect. If this is right, then he is apparently supposing that the emergent structurally determines (in the sense gestured at in note 18) the system’s relational structure.

But what of the further claim that higher-level phenomena “supervene in a way that leaves the micro interactions, per se, unaltered”? It is difficult to be sure of his precise view here, but, as I understand him, Sperry is repudiating a picture of emergence according to which lower-level components that formerly operated invariably in a particular way within a specific local context behave differently in the very same type of local context when the

latter is, in turn, embedded within a complex system exhibiting emergence. Rather, on his view, because an emergent determines (in large part) a relational complex that cannot be adequately described in terms of lower level components and their micro-relations, the micro-physical laws are inadequate for, cannot be applied to, such situations. Hence his remark that “[m]icrodeterminism is thus not so much refuted or falsified, as it is supplemented” (pp.230-1).<sup>33</sup> Of course, the presence and downward control of emergent properties does imply the failure of the overall state of an object or system to be determined by the properties and micro-relations of its parts, and this would ordinarily be taken to be a major component of the thesis of microdeterminism. Finally, this modified wider context impacts the type of causal interactions the entity as a whole has with its surrounding environment.

Sperry briefly remarks on the way in which this approach to emergence is applied in his thesis of emergent consciousness:

Instead of following the usual approaches that tried to inject conscious effects into the already established chain of microcausation, the logical impasse was resolved by leaving the microcausation intact but embedding it within higher brain processes having subjective properties with their own higher-level type of causation, and by which the embedded micro events are thereafter controlled.

. . . excitation of a cortical cell is enjoined into the higher dynamics of passing patterns of cognitive activity. A train of thought with one mental thought evoking another depends throughout on its neurocellular physiology and biochemistry. Nevertheless, like molecules in passing waves in a liquid, the brain cell activity is subject to higher-level dynamics which determine the overall patterns of the neuronal firing, not relative to other events within this particular brain process, but relative to the rest of the organism and its surroundings.

As Sperry readily concedes, his ideas on the nature of consciousness are simply broad research proposals, and rather vague ones at that, in the absence of more developed theories of possible mechanisms of downward control over the electrodynamic current flow within the brain's network circuit systems (p.231). (The previous discussion, however, should have made clear that it is in the nature of the beast that the influence of particular emergent properties qua emergents can be understood only when one first has in hand a fairly thorough grasp of the underlying micro-structures that give rise to it. This is a point that critics of emergentist ideas consistently fail to grasp - perhaps because it hasn't been very well appreciated by emergence theorists either.) But it points in the direction of an exciting alternative approach to the understanding of consciousness that may allow theorists to safely sail between what seems to be the Scylla of substance dualism and the Charybdis of microdeterministic materialism. What I hope to have shown here is that such an approach is both conceptually coherent and empirically viable.<sup>34</sup>

*Indiana University*

---

<sup>1</sup> Vol.I-II (New York: The Humanities Press, 1920).

<sup>2</sup> Alexander draws a distinction between "emergent" and "resultant" qualities in a footnote on p.14.

<sup>3</sup> This is a modified version of David Armstrong's proposal in Universal and Scientific Realism - Vol. II: A Theory of Universals (Cambridge: Cambridge University Press, 1978). (The proposal is made on p.69.) I have introduced a change in order to handle a difficulty that he notes (on pp.69-70) but, to my mind, fails to resolve satisfactorily.

---

<sup>4</sup> Hereafter, for simplicity, I will drop this qualification and speak of a structural property's being identical to the micro-level properties on which it supervenes.

<sup>5</sup> Achim Stephan notes this apparent problem but fails to see how it may be resolved. See p.31 of his "Emergence - A Systematic View on its Historical Facets", in Beckermann, A., Flohr, H., & Kim, J., eds., Emergence or Reduction? - Essays on the Prospects of Nonreductive Physicalism (Berlin: de Gruyter, 1992).

<sup>6</sup> See, e.g., p.62, as well as the footnote on p.14 to which I previously referred.

<sup>7</sup> Contra Brian McLaughlin's suggestion ("The Rise and Fall of British Emergentism", in Beckermann, et. al., 1992, p.66).

<sup>8</sup> Proceedings of the Sixth International Congress of Philosophy (London: Longmans, Green, & Co., 1927), pp.20-33; reprinted in P. Weiner, ed., Readings in the Philosophy of Science (New York: Charles Scribner's Sons, 1953), pp.585-596.

<sup>9</sup> It should be clear that for a property, *P*, of an object, *Q*, to be emergent, it cannot be had by any of *Q*'s ultimate parts. But the reader may wonder why an emergent property could not be shared by one or more of an object's higher-level parts. The reason is that, in such a case, *P* would be emergent with respect to the lowest level in the structure of *Q* exemplifying *P*, but not with respect to the higher levels.

<sup>10</sup> Cf. Armstrong (1978), p.67.

<sup>11</sup> Philosophical Perspectives: Action Theory and Philosophy of Mind 4 (1990), pp.216-226.

---

<sup>12</sup> Kim, "Concepts of Supervenience", Philosophy and Phenomenological Research 45 (Dec.) (1984), pp.153-176 (see p.165); cf. van Cleve (1990), p.220.

<sup>13</sup> We include relational properties so as to circumvent the improper result of spatiotemporally disconnected objects satisfying the definition.

<sup>14</sup> The addition of the bracketed clauses is intended to render his remarks equally applicable to my second definition.

<sup>15</sup> The Mind and Its Place in Nature (London: Routledge & Kegan Paul, 1925), p.59.

<sup>16</sup> Sydney Shoemaker provides a strong defense of this view in "Causality and Properties", reprinted in Shoemaker, ed., Identity, Cause and Mind (Cambridge: Cambridge University Press, 1984) pp.206-233.

<sup>17</sup> "Micro-Determinism and Concepts of Emergence", Philosophy of Science 51 (1984), pp.44-63.

<sup>18</sup> Although one addendum to the definition of supervenience in the text is necessary. For, as Jaegwon Kim notes (1984, pp.166f.), the definition only specifies that the subvening properties entail the supervening property, without implying the failure of the reverse entailment. Hence, it does not capture the asymmetry of dependence that one would intuitively suppose to hold in the case of emergent supervenients (as well as in various other cases). On Kim's way of handling this, we simply add (where appropriate for any given case) that the family of B-properties don't likewise supervene on the A-properties, and thus retain the neutrality of the formal

---

supervenience relation. An alternative would be to build the asymmetry into the concept of supervenience by specifying that the object  $x$ 's having the base property  $b$  at time  $t$  structurally determines  $x$ 's having the supervening property  $a$  at  $t$ , where structural determination at a time is conceived as a species of causation distinct from ordinary efficient causation through time. This, however, raises difficult issues that I haven't the space to explore here. It will suffice for present purposes that we understand the supervenience involved in property emergence to be one-way only.

<sup>19</sup> Compare Klee's (1984) exposition of this idea (his fourth "mark" of emergence) on p.48.

<sup>20</sup> Carl Ginet suggested this objection to me in conversation. There are some similarities to this objection in S. Pepper, "Emergence", *Journal of Philosophy* 23 (1926), pp.241-245, which is criticized by P.E. Meehl and W. Sellars in "The Concept of Emergence", in Feigl, H. et al., eds., *Minnesota Studies in the Philosophy of Science*, I, (Minneapolis: University of Minnesota Press, 1956), pp.239-252. I was led to the response I give in the text by reflecting on Meehl's and Sellars's discussion. (Stephan, 1992, p.45, criticizes Sperry's account of emergent consciousness for essentially the same reason.)

<sup>21</sup> In private discussion, Anil Gupta has objected to this, claiming that since we could give a lawlike account of the quirky scenario without reference to any alleged emergent property (albeit one involving tacked-on disjuncts), nothing is gained in explanatory power by positing such a property. (By contrast, we must quantify

---

over some non-emergent properties if we are to give any lawlike account of phenomena at all.) I deny, however, that one ought to posit a natural property only if (1) doing so is necessary to formulating laws of any kind governing certain phenomena. Most fundamentally, properties are posited (2) to explain differences among various general patterns of events. As it happens, (1) and (2) coincide in non-emergent contexts.

<sup>22</sup> Suggested by Prof. Shoemaker in subsequent discussion.

<sup>23</sup> Franklin Mason took this line in discussion.

<sup>24</sup> McLaughlin (1992), in an otherwise excellent discussion of some of the classical emergentist figures, indulges in rhetoric of this sort.

<sup>25</sup> Although compare R.W. Sperry's claims concerning the "mentalist revolution" in recent neuroscience and psychology. (See, e.g., "Mind-brain interaction: Mentalism, yes; dualism, no" Neuroscience 5 (1980), pp.195-206; "Macro- Versus Micro-Determinism", Philosophy of Science 53 (1986), pp.265-270; "In Defense of Mentalism and Emergent Interaction", Journal of Mind and Behavior 12(2) (1991), pp.221-245.)

<sup>26</sup> "Life's Irreducible Structure", Science 160 (1968), pp.1308-1312.

<sup>27</sup> "A Modified Concept of Consciousness", Psychological Review 76 (1969), pp.532-536. See also, e.g., (1986) and (1991).

<sup>28</sup> A similar claim is found in P. Weiss, "The Living System: Determinism Stratified", in Koestler, A. and Smythies, J.R., eds., Beyond Reductionism (London: Hutchinson, 1968). Klee (1984) provides a very helpful (and critical) discussion of such macro-determinative proposals.

---

<sup>29</sup> And the concept of emergence may well be needed in order to make explicit the recently much-discussed notion of “control hierarchy” in the philosophy of biology, although this remains to be seen. See Uko Zylstra "Living Things as Hierarchically Organized Structures", *Synthese* 91 (1992), pp.111-133 and Marjorie Grene, “Hierarchies in Biology”, *American Scientist* 75 (1987), pp.504-510, for attempts to distinguish biological systems that exhibit hierarchical control from other sorts of hierarchically-arranged systems.

<sup>30</sup> In K.R. Popper and J.C. Eccles, *The Self and its Brain* (New York: Springer International, 1977).

<sup>31</sup> For an interesting discussion of this question in terms of a conceptual framework at least roughly similar to the one developed here, see Dennis M. Senchuk's "Consciousness Naturalized: Supervenience Without Physical Determinism", *American Philosophical Quarterly* 28(1) (1991), pp.37-47.

<sup>32</sup> For other discussions of Sperry's views, see J.J.C. Smart, "Physicalism and Emergence", *Neuroscience* 6 (1981), pp.109-113, and Charles Ripley, "Sperry's Concept of Consciousness", *Inquiry* 27 (1984), pp.399-423, as well as Klee (1984).

<sup>33</sup> Note also in this connection his claim concerning “the inadequacy of science to deal with these spatiotemporal pattern factors other than through their automatic inclusion in laws worked out for the higher levels” (p.236).

<sup>34</sup> I thank Carl Ginet, Sydney Shoemaker, Phil Quinn, and Dean Zimmerman for helpful comments on earlier drafts of this paper. I

---

read the paper to the philosophy departments at Indiana University and Purdue University, and received a number of helpful, trenchant suggestions from those audiences (and in particular, from Anil Gupta, William Rowe, and Franklin Mason). Finally, I have been aided in thinking about this topic by the very recent collection of new essays on emergence in Beckermann, et. al. (1992), which appeared subsequent to the initial drafts of this paper. My account differs in important respects from each of those proposed therein, though I have not tried to point out such differences here.